

Doctoral Programme in Plant Sciences
Faculty of Agriculture and Forestry
University of Helsinki

GENOMIC PREDICTION IN PRACTICAL BREEDING PROGRAM: A CASE STUDY IN OAT AND BARLEY

Hanna Haikka

DOCTORAL THESIS

to be presented for public examination with the permission of the Faculty of
Agriculture and Forestry of the University of Helsinki, in Raisio hall, Forest Sciences
Building, on the 23rd of April, 2021 at 13 o'clock.

Supervisor	Adjunct Prof. Outi Manninen, Boreal Plant Breeding Ltd.
	Prof. Teemu Teeri, Department of Agriculture, University of Helsinki
Follow-up group	Prof. Mikko Sillapää, Research Unit of Mathematical Sciences, University of Oulu
	Dr. Merja Veteläinen, Boreal Plant Breeding Ltd.
Reviewers	Prof. Gunter Backes, Faculty of Organic Agricultural Sciences, University of Kassel
	Prof. Rodomiro Ortiz, Department of Plant Breeding, Swedish University of Agricultural Sciences
Opponent	Prof. Hermann Büstmayr, Department of Agrobiotechnology, University of Natural Resources and Life Sciences, Vienna
Custos	Prof. Teemu Teeri, Department of Agriculture, University of Helsinki

The Faculty of Agriculture and Forestry uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

ISSN 2342-5423 (print)
 ISSN 2342-5431 (online)
 ISBN 978-951-51-6835-1 (pbk.)
 ISBN 978-951-51-6836-8 (PDF)
<https://ethesis.helsinki.fi>

Unigrafia Oy
 Helsinki 2020

ABSTRACT

The aim of this study was to detect usability of genomic prediction for different breeding dilemmas. In order to achieve this aim, breeding data sets from oat and barley were used in the study. The studied lines were genotyped with genome-wide markers. Meanwhile, phenotypes were collected from multiple years and locations of historical breeding data. Together, the data of the line genotypic and phenotypic information formed the training population used in the analysis. The separate studies concerned genomic prediction, genome-wide association study (GWAS) and analysis on genotype by environment (GE) interaction. The studies had in common that they present 'difficult' topics within the breeding process.

The original publication I concentrated on improving grain yield prediction for oat and barley. Grain yield presents one of the most important traits in breeding, but has low predictability due to low heritability. The prediction of genomic estimated breeding values (GEBVs) was improved by using multi-trait prediction. For this purpose, grain yield was predicted simultaneously with correlated traits. In addition, benefit of trait-assisted prediction was examined. In conclusion for oat and barley, prediction of grain yield was improved by 4% and 9% with multi-trait prediction, and by 9-14% and 11-28% with trait-assisted prediction compared to prediction of grain yield alone, respectively.

The original publication II focused on Fusarium head blight (FHB) resistance in oat. FHB resistance is a troublesome trait to breed, since the disease cannot be reliably scored visually, but extensive laboratory analysis is needed to obtain resistance phenotypes. In addition, FHB resistance consists of multiple components. In the study, the correlations between FHB resistance related traits were high. Much lower correlations were seen between FHB resistance related and agronomic traits. No significant associations between FHB related traits and genetic markers were discovered with reasonable correction of population structure and genetic relationship between the studied oat lines. For this reason, using genome-wide marker information to promote resistance breeding should be done solely with genomic selection (GS), where all the marker effects are used to enrich resistance alleles within the breeding population.

The original publication III explored the extent of GE interaction within breeding data sets of oat and barley. At first, the genetic correlations between trial locations within year were calculated and used to compute mean across the years. The correlations suggested that data set of oat was not as sufficient as the data set for barley to explore the quantity of GE interaction. The second step of the analysis contained genomic prediction with six different models. The prediction models contained effects due to lines, genetic, environmental covariates, GE interaction and genotype by environmental covariates

interaction. The prediction accuracy was increased for both crops when GE interaction was added into the prediction model. The results from the analysis imply that GE interaction exists within the breeding data sets, and should be taken into account upon prediction.

All of the conducted studies proved the usability of genomic prediction in solving principal questions in the breeding process. The studies improved prediction of central traits simultaneously enabling the prediction in the early breeding generations, and showed the significance of GE interaction, and most of all, showed that historical breeding data can be used to predict the important traits. These studies present tools for practical breeding in order to meet the demand to accelerate crop improvement.

TIIVISTELMÄ

Yksi jalostuksen keskeinen käsite on jalostusarvo. Se määrittää kuinka hyvä lajikekandidaatti on periyttämään haluttuja ominaisuuksia. Perinteisesti jalostusarvo saadaan selville lajikkeen jälkeläisiä tarkastelemalla. Jalostusarvon genomista ennustamista varten tarvitaan lajikekandidaatin perimä- ja ominaisuustietoja sen sukulaisista. Genomisella ennustamisella jalostusarvo saadaan selville nopeammin eikä jälkeläisten tarkastelua enää tarvita. Genomista ennustamista on hyödynnetty eläinjalostuksessa jo pidemmän aikaa. Eläimen jalostusarvo saadaan selville jo syntymässä eikä sen tuottamien jälkeläisten tuloksia tarvitse enää odottaa. Kasvinjalostuksessa genomisen ennustamisen mahdollisuuksia vasta tutkitaan laajemmin. Tämän väitöskirjan ensisijainen tavoite oli selvittää genomisten ennusteiden käytettävyyttä kasvinjalostuksessa. Kasvinjalostuksessa tuotetaan joka vuosi paljon uutta tietoa. Jalostusprosessissa lajikekandidaatteja testataan erilaisissa ympäristöissä useina vuosina. Tämän väitöskirjan tutkimusaineisto koostui kauran ja ohran jalostusohjelmissa jo valmiiksi kerätyistä tiedoista. Tutkimukset keskittyivät jalostusprosessin vaativiin aiheisiin, kuten sadon ennustamiseen, punahomeen kestävyys- sekä genotyyppin ja ympäristön yhdysvaikutukseen.

Väitöksen ensimmäisessä julkaisussa tarkasteltiin kauran ja ohran sadon genomista ennustamista. Satotason nostaminen on yksi jalostuksen tärkeimmistä tavoitteista, mutta satoisuuden ennustettavuus on heikko alhaisen periytyvyyden vuoksi. Sadon genomisten jalostusarvojen (GEBV) ennustamista parannettiin käyttämällä monimuuttujamenetelmiä, joissa hyödynnettiin sadon kanssa yhteydessä olevia ominaisuuksia, kuten kasvuaikaa ja valkuaispitoisuutta. Kauralla onnistuttiin parantamaan sadon ennustamiskykyä 4% ja ohralla 9% monimuuttujamenetelmän avulla. Lisäksi tutkittiin ominaisuusavusteisen ennustamisen hyötyjä, missä jalostuslinjalla ei ole vielä satotuloksia, mutta jo olemassa olevia kasvuaika- ja valkuaispitoisuustietoja käytettiin ennustamishetkellä. Kauralla ominaisuusavusteinen ennustaminen paransi ennustamiskykyä 9-14% ja ohralla 11-28% verrattuna sadon ennustamiseen yksittäisenä ominaisuutena. Tulokset olivat merkittäviä parannuksia ennustamiskykyyn ja ennen kaikkea tärkeä löytö oli, että ennustamiskyky ei huonontunut monimuuttujamenetelmiä käytettäessä.

Väitöksen toisessa julkaisussa keskityttiin punahomeen kestävyys kauralla. Punahomeet (*Fusarium*-sienet) voivat muodostaa viljojen jyviin myrkyllisiä yhdisteitä (hometoksiineja), joille on määritetty EU:ssa raja-arvot. Saastuneet viljaerät aiheuttavat taloudellisia tappioita viljelijöille. Punahomeen kestävyys kauralla on haastava ominaisuus jalostaa, koska tautia ei voida luotettavasti havaita pellolla vaan kestävyys selvittäminen vaatii laboratorioanalyysia. Lisäksi punahomeen kestävyys ei ole vain yksi

ominaisuus vaan koostuu useammasta toisiinsa liittyneestä ominaisuudesta. Tutkimuksessa tehdyssä assosiaatiokartoituksessa ei löytynyt punahomeen kestävyteen vaikuttavia geenialueita. Tuloksista voitiin päätellä, että kestävyden nostamiseksi olisi parempi käyttää genomisia ennusteita, joiden avulla ei hyödynnetä niinkään yksittäisiä kestävyteen liittyviä geenialueita vaan rikastetaan pienivaikuttaisia kestävyttä parantavia geenejä koko genomissa.

Väitöksen kolmannessa julkaisussa tutkittiin genotyyppin ja ympäristön yhdysvaikutuksen laajuutta kauran ja ohran jalostusohjelmissa. Yhdysvaikutus ilmenee, kun eri testauspaikoilla, tai eri vuosina, tutkittujen lajikekandidaattien paremmuusjärjestys vaihtelee huomattavasti. Sama lajikekandidaatti ei ole paras jokaisella testauspaikalla. Tutkimuksen tulokset viittasivat siihen, että kauran tutkimusaineisto ei ollut yhtä riittävä aiheen tutkimiseen kuin ohralla. Tutkimuksessa myös ennustettiin satoa kuudella tilastollisella mallilla, joiden paremmuutta vertailtiin. Edistyneemmissä malleissa hyödynnettiin yhdysvaikutusta ja ympäristöä kuvaavia muuttujia, kuten testauspaikkojen säätietoja. Ennustamiskyky parani molemmilla viljoilla, kun yhdysvaikutus lisättiin ennustemalliin. Tutkimuksen tulokset viittasivat siihen, että yhdysvaikutusta esiintyy molemmissa jalostusohjelmissa, ja se tulisi ottaa huomioon satoa ennustettaessa.

Väitöksessä tehdyt tutkimukset osoittivat genomisen ennustamisen käyttökelpoisuuden jalostuksessa. Tutkimuksissa onnistuttiin parantamaan keskeisten jalostusominaisuuksien ennustamiskykyä, mikä voisi mahdollistaa jalostuksen nopeutumisen ja tarkentumisen. Lisäksi tutkimuksessa osoitettiin yhdysvaikutuksen merkitys jalostusohjelmissa, ja ennen kaikkea todettiin, että historiallista jalostusohjelmassa kerättyä tietoa voidaan käyttää tärkeiden ominaisuuksien genomiseen ennustamiseen. Tutkimukset antavat käytännöllisiä työkaluja jalostukseen. Nämä työkalut ovat hyvin arvokkaita, kun halutaan vastata ruoantuotannon lisääntyviin haasteisiin, parantaa viljelyn tuottavuutta sekä tehostaa jalostusprosessia.

ACKNOWLEDGEMENTS

This thesis was carried out thanks to many participants on the way. First of all, I would like to thank Boreal Plant Breeding Ltd. for given me access to their true breeding data sets to play with and use in publications. I cannot thank enough professor Mark Sorrells and his lab from Cornell University for given me tools and courage to learn as much as I could from genomic selection. My two-semester visit to Cornell was enabled by former Finnish Koulutusrahasto (current Employment Fund) and Boreal Plant Breeding Ltd. As well, the work was financed by Agronomiliitto (Suomi kasvaa ruoasta -scholarship paid by Henrik and Ellen Tornbergin foundation), a travel grant from University of Helsinki in order to revisit Cornell and writing retriets at University of Helsinki reaserch stations. I also would like to thank all the numerous seminar organizers and especially, CIMMYT, The International Maize and Wheat Improvement Center and Jose Crossa, for given me guidance on my study path.

Special thanks to wonderful supervisors Outi Manninen and Teemu Teeri. Outi was always challenging me and pushing me forward, and without her I would not be here. Teemu understood, supported and made the communication effortless for this industrial PhD. Especially, thanks for keeping up the timetable during structural changes at the University of Helsinki. My follow-up group, Mikko Sillanpää and Merja Veteläinen, I would like to thank for given me their support and for very valuable and guiding comments on the way. Pre-reviewers, Rodomiro Ortiz and Gunter Backes, are thanked for taking this task and providing valuable feedback. Finally, my opponent, Hermann Bürstmayr, I wish to thank for given me this opportunity and for patience during unexpected times due to Covid-19.

The original publications would not have been published without the help from statistical encyclopedia Timo Knürr, animal research professors from Luke Esa Mäntysaari and Ismo Strandén, Boreal plant breeders, Leena Pietilä, Mika Isolahti and Esa Teperi, co-workers from Luke Marja Jalli and Juho Hautsalo and staff from Luke involved in Fusarium work, and most of all, all the staff from Boreal Plant Breeding involved in oat and barley breeding programs. Thank you all!

Of course, I want to thank my family and friends for their support and understanding why and where I have been hiding for a couple of years. Finally, with tears, I wish to thank my husband Aleksi Friman who supported me all the way, specially during the darkest times, when I felt that I could not do this. Now, we can say that this project was a piece of cake, “helppo nakki” :) For our unborn child, your little kicks pushed this book forward and reminded me about other things in life and life after this effort.

October, 2020

CONTENTS

Abstract.....	3
Tiivistelmä	5
Acknowledgements.....	7
Contents.....	8
List of original publications.....	10
Abbreviations.....	11
1 Introduction.....	12
1.1 Crop plants	12
1.1.1 Oat (<i>Avena sativa</i> L.).....	12
1.1.1.1 Genome and domestication	12
1.1.1.2 Uses and cultivation	13
1.1.2 Barley (<i>Hordeum vulgare</i> L.)	14
1.1.2.1 Genome and domestication.....	14
1.1.2.2 Uses and cultivation	15
1.2 Introduction to plant breeding	15
1.2.1 Quantitative and qualitative traits.....	17
1.2.2 Selection	17
1.2.3 Genotype by Environment interaction	19
1.3 Genomics in plant breeding.....	20
1.3.1 Genetic information	21
1.3.2 Genomic selection	22
1.3.2.1 Methods of genomic prediction	22
1.3.2.2 Accuracy of genomic prediction	24
1.3.2.3 Factors affecting prediction accuracy.....	25

1.3.2.4	Applications of GS	26
1.3.3	Genome-wide association study	26
1.3.3.1	Methods of GWAS.....	27
1.3.3.2	Factors affecting GWAS.....	27
2	Aims of the study	30
3	Materials and methods.....	31
3.1	Breeding of self-pollinated crops.....	31
3.2	Breeding data sets	32
3.3	Genotypes and Population structure.....	32
3.4	Variance component estimation and genomic prediction	33
3.5	Validation of prediction	33
3.6	Association study	33
4	Results and discussion	35
4.1	Improving prediction of grain yield	35
4.2	Genomics of FHB related resistance traits.....	37
4.3	GE interaction within breeding populations.....	38
4.4	Applying genomic selection to breeding programs.....	39
5	Conclusions and future prospects	43
	References	45
	Original Publications.....	61

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications, which are referred in the text with their roman numerals.

- I Haikka, H., Knürr, T., Manninen, O., Pietilä, L., Isolahti, M., Teperi, E., Mäntysaari, E. A. and Strandén, I. Genomic prediction of grain yield in commercial Finnish oat (*Avena sativa* L.) and barley (*Hordeum vulgare* L.) breeding programmes (Plant Breeding, 2020, 139:550–561, doi: 10.1111/pbr.12807)
- II Haikka, H., Manninen, O., Hautsalo, J., Pietilä, L., Jalli, M. and Veteläinen, M. Genome-wide association study and genomic prediction for *Fusarium graminearum* resistance traits in Nordic oat (*Avena sativa* L.) (Agronomy, 2020, 10, 174, doi: 10.3390/agronomy 10020174)
- III Haikka, H., Manninen, O., Pietilä, L., Isolahti, M., Teperi, E. and Knürr, T. Genotype by environment interaction and environmental covariates in Finnish commercial oat (*Avena sativa* L.) and barley (*Hordeum vulgare* L.) breeding programs (manuscript)

The original publications have been reprinted with the kind permission of their copyright holders.

Contribution of authors can be detailed as follow:

- I Together HH, TK, OM, EM and IS designed the study and participated in the interpretation of results. HH performed the data analysis, while TK advised. HH drafted the manuscript. All authors critically revised the manuscript. The data was collected by LP and MI. ET, LP and MI performed the pre-study data analysis.
- II The study was designed by OM and HH. Both participated in the interpretation of results. HH performed the data analysis and drafted the manuscript. All authors critically revised the manuscript. Data was collected by JH, MJ and LP.
- III Together HH, OM and TK designed the study and participated in the interpretation of results. HH performed the data analysis and drafted the manuscript. All authors critically revised the manuscript. Data was collected by MI and LP. ET, LP and MI performed the pre-study data analysis.

ABBREVIATIONS

AMMI	Additive Main effects and Multiplicative Interaction
BLUE	Best Linear Unbiased Estimation
BLUP	Best Linear Unbiased Prediction
DH	Doubled-Haploid
DON	Deoxynivalenol
EBV	Estimated Breeding Value
e.g.	exempli gratia
etc.	et cetera
FHB	Fusarium Head Blight
FIK	<i>Fusarium</i> Infected Kernels
GBLUP	Genomic Best Linear Unbiased Prediction
GC	Germination Capacity
GE	Genotype by Environment (interaction)
GEBV	Genomic Estimated Breeding Values
GS	Genomic Selection
GWAS	Genome-Wide Association Study
i.e.	id est
LASSO	Least Absolute Shrinkage and Selection Operator
LD	Linkage Disequilibrium
Luke	Natural Resources Institute Finland
MAGIC	Multi-parent Advanced Generation Inter-Cross
MAS	Marker Assisted Selection
MSE	Mean-Square Error
NAM	Nested Association Mapping
PCA	Principal Component Analysis
PLS	Partial Least Squares
qFUSG	<i>Fusarium graminearum</i> DNA (relative to oat DNA) content
QTL	Quantitative Trait Loci
RKHS	Reproducing Kernel Hilbert Space
rrBLUP	Ridge Regression Best Linear Unbiased Prediction
SNP	Single Nucleotide Polymorphism
SREG	Site REGression
SSD	Single Seed Descent
TBV	True Breeding Value
VYR	the Finnish Cereal Committee

1 INTRODUCTION

1.1 CROP PLANTS

Cereal crop plants have evolved from wild species during thousands of years and first bread-like products are dated 14,400 years ago (Arranz-Otaegui et al. 2018). During domestication process plants were selected intentionally and unintentionally to meet the human needs, and so-called domestication syndrome traits (Hammer 1984), like uniform ripening, erect growth habit and increased seed size and number, differentiated crop plants from their wild progenitors (Zohary et al. 2012). These traits pronounced first forms of selection, which progressed for several thousands of years improving the crops for humans needs. During recent history, selection within local strains started after 1850s (Thomas 1995). Modern crop improvement with artificial pollination is a fairly recent practice, and was initiated mostly after Mendel's article "Experiments on Plant Hybrids" (1866) (Voss-Fels et al. 2019).

Only a few crop plants feed the world today. Measured by harvest area 25 crops use 83.7% of the total cultivated area (FAO 2017a). Wheat, maize, rice, soybean and barley are the most cultivated species while oat is the 25th. Despite regional variability, overall yields of these major crops may be reduced due to climate change in the long run (Porter et al. 2014). Improving crop plants is the main task of plant breeding. Most important goals for breeding are more yield per measured area of production, more resilient crops and better quality for human use (Porter et al. 2014, Mickelbart et al. 2015). In order to achieve these goals, plant breeders are improving their methods, accelerating the breeding process and closely following the progress in plant research (Voss-Fels et al. 2019, Atlin et al. 2017, Forster 2014, Lenaerts et al. 2019). Crop improvement is needed to cover the growing demand of food, which has been estimated to increase 50% by the year 2050 (FAO 2017b). This study focuses on improving breeding methods for two cereal crops: oat and barley.

1.1.1 OAT (*Avena sativa* L.)

1.1.1.1 Genome and domestication

Oats belong to plant family of Poaceae, but diverge from other small-grained cereals and belong to Aveneae tribe. Cycles of interspecific hybridizations and polyploidizations have formed the cultivated oat (*A. sativa*), which carries three genomes designated AA, CC and DD each containing seven chromosomes (Rajhathy et al. 1974). The ancestral species for oat genomes are not as certain as for the three genomes of wheat (Fu 2018). Studies show that A and D genome are more similar and C genome is divergent from them (Jellen

et al. 1994, Peng et al. 2010, Latta et al. 2019). Primary chromosome pairing is disomic, but nonhomologous pairing is common (Chaffin et al. 2016). Irregularities in chromosome pairing, size of the genome, translocations and lack of sequence data have hindered formation of high-density consensus maps (Chaffin et al. 2016, Oliver et al. 2011, 2013). *A. sativa* genome has a lot of repetitive DNA and its haploid genome size is estimated to be 12.6 Gbp (Yan et al. 2016). At least two major translocations have been described, one in chromosomes designated as 7C and 17A (Jellen and Beard 2000), and another in chromosomes 3C and 14D (Jellen et al. 1997). For now, there is no fully sequenced, publicly available reference genome for *A. sativa*. The availability of the reference would result in more precise alignments of genetic markers used in various analysis.

Oats have most likely evolved as a weed for wheat and barley four to five thousand years ago (Valentine et al. 2011). Oats have two hypothetical centers of domestication, the Near East and the Western Mediterranean (Jellen and Beard 2000). From the Near East, the common oat (*A. sativa*) spread to Europe in the late Bronze Age (Valentine et al. 2011) and to North America in the 16th century (Coffman 1977). The common oat has a wide range of relatives in the *Avena* genus. Unlike wheat, oat has wild forms of hexaploid species, like *A. sterilis*, which has been proposed as a potential ancestor (Coffman 1946, Li et al. 2000). While common oat in the North America contained germplasm from both *A. sativa* and *A. byzantina*, the European germplasm has been described to have narrower genetic base than germplasm in North America (Valentine et al. 2011).

1.1.1.2 Uses and cultivation

Oats are used for human consumption, animal feed and industrial applications (Marshall et al. 2013). Oat production worldwide was 25.9 million tons in 2017 (FAO 2017a). With production over million tons each, Russia, Canada, Australia, Poland, China and Finland together deliver 59% of the world production. Most of the production is for human consumption, breakfast cereals and porridge, but also a broader use of oat-based products (Marshall et al. 2013). Oat's health benefits come mainly from high soluble β -glucan content (Lee et al. 1997), but oats are also healthy protein source, with rich profile in vitamins, minerals and antioxidants, like α -tocotrienols, α -tocopherols and avenanthramides (Peterson 2001). Oats have traditionally been used for grain feed, where the benefit is in high oil content compared with other cereals (Welch 2011). Besides grain feed, oats are used worldwide as fodder in grazing, hay and silage, as well. The cosmetic and pharmaceutical industry increasingly demand oat fractionation products, like β -glucan extracts and oil derivatives. These are used both in human health products and cosmetics (Marshall et al. 2013).

Marshall et al. (2013) stated that oat has a “low demand in nitrogen, low susceptibility to cereal diseases and high competitiveness with weeds”, which

makes it a relatively low input crop compared to other cereals and suitable for organic production. However, in the main production areas the need to use the sufficient inputs in cultivation has been realised in order to produce high yields with good quality. Oat's major challenges for better production are lodging, yield improvement and susceptibility to both crown rust (*Puccinia coronata* f.sp. *avenae*) and Fusarium head blight (Marshall et al. 2013). No efficient dwarfing genes have been used in oat breeding, contrary to wheat. While yield level of oats has risen, the increase has been less than in wheat and barley (Marshall et al. 2013, Öfversten et al. 2004). Crown rust presents the most devastating disease for oats (Simons 1985) and mycotoxins produced by *Fusarium* fungus impact the quality of this crop (Marshall et al. 2013).

1.1.2 BARLEY (*Hordeum vulgare* L.)

1.1.2.1 Genome and domestication

Barley belongs to the family of Poaceae and shares the tribe Triticeae with wheat. The common barley (*H. vulgare*) is a diploid with seven chromosomes. Barley exhibits mechanisms to prevent cross-pollination, such as fertilization occurring before heading (Alqudah and Schnurbusch 2017). Therefore, due to the simple disomic inheritance, diploid genome and sharing the same tribe, barley has been used as a model genome for more complex wheat. The use as a model was reinforced when barley genome was fully sequenced (Mayer et al. 2012). The size of the sequenced haploid genome is relatively large (5.1 Gb).

Barley is one of the oldest cereals. During Neolithic revolution 13000-10000 years ago in Near East, first agricultural societies were formed (Purugganan and Fuller 2009). Even before that people used wild cereals, which were in the process of transforming from wild to domesticated, along with seeds and nuts (Wendorf et al. 1979, Kislev et al. 1992). Two-row barley appeared earlier than six-row types (Zohary et al. 2012). Gradually the common barley spread to Europe and Asia 8000 years ago (Zohary et al. 2012), and to northern Europe 6000 years ago (Briggs 1978). The *Hordeum* genus contains species with high biodiversity. However, morphologically they are relatively similar, the main difference being the sterility (two-row barley) or fertility (six-row barley) of lateral florets (Briggs 1978). The ancestor of the common barley is *H. vulgare* ssp. *spontaneum*, which belongs to the primary gene pool according to the classification done by Harlan and de Wet (1971). Common barley can be successfully hybridized with its ancestor. Therefore, the ancestor species can be used as a gene repository for alleles that have already disappeared from the common barley. Produced progeny is viable and produces seed.

1.1.2.2 Uses and cultivation

Barley is mainly produced for feed, malt, alcohol production and human consumption. Production worldwide in 2017 was 147 million tons (FAO 2017a). The five most producing countries are Russia, Australia, Germany, France and Ukraine. The main portion of barley is produced for feed and the second largest user is malting industry (Newton et al. 2011). Human consumption is relatively small, even though barley has high β -glucan level (Newman et al. 1989). The most common storage protein, hordein, contains largely unessential amino acids, proline and glutamine, which makes the nutritional quality of barley modest (Doll 1983). Bowman et al. (1996) described the valuable characteristics of barley being high energy feed for cattle.

Barley is a versatile crop and can be cultivated in a wide range of environments from tropics to high altitudes and latitudes (Paulitz and Steffenson 2011), which has been the main driver of its use as a food crop. In principal, barley is more productive and stable than wheat (Newton et al. 2011). Even though barley is a relatively resilient crop, the most important improvements are required in abiotic and biotic stress tolerance traits, which are listed extensively in Newton et al. (2011). In order to maintain resilience in barley germplasm breeders should avoid loss of biodiversity.

1.2 INTRODUCTION TO PLANT BREEDING

Plant breeding involves a specialized field of science, aiming to improve crop plants for human benefit. The main task of breeding is to develop new, improved cultivars from a breeding population through a breeding process, but simultaneously breeding uses and combines many fields of science to improve genetic potential of crops and produces scientific discoveries (Bernardo 2010). Breeding progress can be measured and guided by the breeder's equation (1) originating from Lush (1943):

$$(1) \quad g = \frac{\sigma i r}{L}$$

where genetic gain (g) is defined by multiplying genetic variation within the population (σ) with selection intensity (i) and selection accuracy (r). Dividing of the equation with time (L) was introduced by Eberhart (1970).

Plant breeding starts by creating a breeding population. This involves selection and crossing the best available parents to form a progeny population. In principal, breeding populations should have both a high mean and a large genetic variance for the traits of interest. High mean secures superiority of the breeding population and large variance contributes potential for genetic improvement (Bernardo 2010). As favorable alleles are selected, they will by time become fixed, and genetic variation within a population is decreased.

According to breeder's equation, without genetic variance, genetic gain within a breeding population is nil.

The breeding process of self-pollinated cereal crops results in superior inbreds and can be roughly divided into three parts. First, new genetic variation is created through recombination by crossing multiple pairs of parents to form a breeding population (referred to later as a breeding cohort). Second, the created F_1 population is highly heterozygous and inbreeding is needed in order to have a breeding lines with stable characteristics. Inbreds can be developed with multiple methods. A comprehensive review can be found in the literature (Thomas et al. 2003). In pedigree selection (Briggs 1978) repeated selection rounds promote selfing. Selection is done by sowing biparental families as rows into field and selfed seed of selected individuals is used for planting the next generation of head rows. Traits, which can be reliably measured from an individual plant, are selected during the process. The breeding population, in this case, contains biparental populations and sub-biparentals (family rows) and, therefore, selection can be executed in both within and between families. Differing from pedigree selection, in single seed descent (SSD) only one seed from each plant is used for the next generation (Brim 1966) and commonly greenhouse is used in order to accelerate the production of selfing generations. Otherwise principles between the two methods are similar, except sub-biparentals are missing in SSD and possible selection is practiced between families only. Generally, selection starts after selfing generations, but marker-assisted selection (MAS) could potentially be used during the process (Collard and Mackill 2008). The defined unit of a family is replaced around F_7 generation, where visual variation within a family is minor and the unit of selection is called a breeding line from there onwards (Forster 2014). The fastest method to overcome heterozygosity is to make doubled-haploids (DH) with tissue culture technique (Devaux and Kasha 2009, da Silva Dias 2003). In DH method, haploid microspores are used to reach perfect homozygosity within one generation, which saves time compared to the methods described above. However, only one round of recombination occurs if DH plants are created from F_1 generation which limits the reshuffling of genes in the breeding population. Third part of the breeding process focuses on testing and selection, where best cultivar candidates are selected from the breeding population through intensive testing in multiple environments and years. In early generations the number of cultivar candidates is large but quantity of seed and thus amount of testing is limited. The amount of lines decreases with selection during the process and gradually more precision is reached. At the end of the process accurate information for cultivar candidates is received and the best cultivar candidates are promoted to official testing of their cultivation value.

The breeding process is always limited with time and resources. As mentioned above, the amount of lines in the beginning of the process is large, and has to be reduced before extensive grain yield or quality trait testing can begin, due to extensive costs of field trials and quality analysis. Time is an

important factor, when efficiency of the process is estimated. One measure of time in the breeding process is the breeding cycle, which is defined as the time from a line is used as a parent until one of its progenies is used as a parent. It takes multiple years to find the optimal crossing parents. In order to evaluate parents, the seed of the breeding line is increased to adequate amount, which takes years. After seed multiplication, adequate evaluation and testing should be done in order to have a reliable judgement of the breeding line. Eventually, to achieve enhanced genetic gain, the breeding cycle should be as short as possible with adequate selection accuracy (Cobb et al. 2018).

Three important issues affecting the breeding process of self-pollinated cereal crops are explained in the following chapters.

1.2.1 QUANTITATIVE AND QUALITATIVE TRAITS

The genetic architecture of target traits for breeding differs greatly. Some traits are influenced by genes from one to a few loci and others by hundreds of loci. In breeding context, the traits are divided into qualitative or quantitative traits, where the definition and division point are vague. Qualitative traits are controlled by a few major genes, sometimes by even only one gene with a large effect. Generally, qualitative traits are relatively easy to characterize by the Mendelian rules of inheritance. The traits can often be indicated by only a limited number of levels or distinct categories. On the contrary, quantitative traits are controlled by a large number of minor genes with small effects which accumulate into so-called polygenic effects and create a continuum in the distribution of the trait. One key aspect of quantitative traits is that environment may have a greater influence on them than on qualitative traits (Bernardo 2010). Grain yield, a quantitative trait, represents the most important trait in commercial breeding programs. However, qualitative traits also have great significance in crop improvement, especially in resistance breeding. As an example of a major gene giving a great economical value, is the powdery mildew resistance gene, *mlo* in barley (Jørgensen 1992, Büschges et al. 1997).

1.2.2 SELECTION

The traits of interest are exposed to selection during the breeding process. Two key concepts have a large influence on selection: heritability of the trait and correlation between traits. Broad sense heritability of a trait is the portion of the overall variation (i.e. phenotypic variance) explained by the genotype (i.e. genetic variance divided by phenotypic variance). Genetic variance includes variance due to dominance and epistasis in addition to additive genetic variance. Narrow sense heritability describes the portion of phenotypic variance explained by additive genetic variance alone. Additive effects describe allele substitution and additive genetic variance is the portion of genetic variance which is inherited to the next generation. Heritability is affected by

the genetic background of the trait, but it also depends on the population under evaluation, the environment for testing and the method used for measuring the trait. If the population contains genetically diverse collection of plant lines, the portion of genetic variation is higher and therefore heritability is also higher compared to a population with a narrower genetic base. The test environment, like an uneven experimental field, may promote phenotypic variability, which cannot always be corrected with experimental designs. This induces noise to the measurement and lowers heritability. The measurement practice has an impact on heritability, since some methods are more precise than others. The traits with higher heritability should be used in selection in the early breeding process, where limited size of the testing unit and low number of records add uncertainty to selection.

Usually in practical plant breeding multiple traits are selected simultaneously. The correlation between traits can be described by means of phenotypic correlation or additive genetic correlation. In phenotypic correlation, traits are correlated, because of genetic and nongenetic causes, like crop management. Additive genetic correlation is either due to linkage, where loci affecting different traits are within close range in the same chromosome, or pleiotropy, where different traits are controlled by the same loci. Correlation between traits can promote selection, when correlation is positive and breeding aims at increasing both traits. In the case of negative correlation, selection of traits simultaneously is hindered if selection aims at increasing both traits, but selection is favorable in cases when breeding for opposite directions is aimed at. In all cases, an index can be formed, where traits of interest are combined and each of them borrows information from others based on the level of correlation shared. The index is used in selection instead of original traits (Hazel et al. 1994). Such a well-known index can be defined for yield and quality traits, which often have a negative correlation. Another way to use correlation is indirect selection, where the trait with low heritability is selected indirectly via highly correlated trait, which has higher heritability compared to original trait (Falconer and Mackay 1996).

Two distinguished parts of selection can be recognized within a breeding process: selection across breeding cycle and selection within breeding cycle. In the first case, in order to create variation, the best possible parental lines are selected from the breeding cohort to increase the genetic value in the next generation (i.e. parental selection). In this case, effects, which are inherited to the progenies (i.e. additive effects), are computed as an estimated breeding value (EBV). However, often information of breeding values is not available and breeders are forced to use phenotypic selection. In the second type of selection, cultivar candidate lines are selected for the next phase of field testing. The genetic value of a cultivar candidate line is influenced by both additive and nonadditive effects. In case of self-pollinated crops nonadditive effects are caused by epistasis. Therefore, instead of a breeding value, the total genetic value, indicating the commercial value of the cultivar candidate, should be estimated (Crossa et al. 2017).

1.2.3 GENOTYPE BY ENVIRONMENT INTERACTION

In the breeding process both traits and selection are influenced by genotype by environment (GE) interaction. GE interaction is present in human (Baye et al. 2011) and animal populations (Falconer and Mackay 1996) as well, but repeated measurements of inbred lines in different environments makes it easier to detect in plant breeding process (Falconer and Mackay 1996). GE interaction refers to the difference in genotype responses in different environments. The best cultivar candidate tested in environment 1 may not be the best in environment 2. Factors affecting this ranking change are genetically and environmentally driven. First of all, the level of GE interaction depends on the crop and the breeding population in question (Burgueño et al. 2011, Yan et al. 2016, Pauw et al. 1981). Second, environmental variables, such as soil type, climatic factors (e.g. precipitation and temperature), the amount and quality of sunlight and pests, pathogens and weeds present (Comstock and Moll 1963), give an environment its specific characteristics, to which plants respond. Differing response of genotypes can lead to different patterns of GE interaction. Genotypes may not have a response and rank in a same way for two different environments. Genotypes may have a different level of response but rank does not change. Rank between genotypes can be opposite and cross-over interaction is observed (Ouyang et al. 1995). Extent of the response and complexity of GE interaction dictates the requirement for field testing in multiple years and locations. The testing locations and years of a breeding program should catch variable environments, so that many of the possible GE interactions of the cultivation target area would be revealed during selection process.

GE interaction can be treated in many different ways. Bernardo (2010) listed three different approaches of how to cope with GE interaction in the plant breeding context. First approach is to ignore it. Even if GE interaction exists, it is coped by testing cultivar candidates in a vast set of environments, and superior cultivar is the one with highest mean across environments. In the second approach, GE is reduced. The target environment is divided into small enough sets, in order to reduce the significance of GE interaction. Third approach seeks to exploit GE. This means that cultivars are bred for specific environments, and GE interaction is studied, by stability analysis or multiplicative models in order to take advantage of the interaction rather than ignoring or reducing it. Stability analysis requires a measure that separates the environments from each other (Bernardo, 2010). Ideally, this means that environmental factors are available, such as climatic factors, biotic or abiotic stresses. If these are not available, mean yield in each environment minus the overall mean yield can be used to separate and scale the environments (Eberhart and Russell 1966, Bernardo 2010). Stability analysis can be executed with joint-regression analysis (Yates and Cochran 1938, Eberhart and Russell 1966, Finlay and Wilkinson 1963, Lin et al. 1986), where a linear model is fitted for genotypes and regression coefficient and variance of the deviation of a genotype in an environment from the fitted value are used as descriptive

statistics. The regression coefficient describes the response of a genotype and the variance how much of the variation is explained by the coefficient (Bernardo 2010). Stability analysis is a simple form of a multiplicative model, as one term of environmental factor is used, whereas in multiplicative analysis, multiple terms can be used simultaneously (Williams 1952). Additive main effects and multiplicative interaction (AMMI) (Gauch 1988, Gauch 2013) and site regression (SREG) (Crossa and Cornelius 1997, Crossa et al. 2002) models are used to quantify and illustrate GE interaction with biplot images. The described models are mostly used as fixed effect models without variance-covariance structure, but including information on genetic relationships between individuals with these variance-covariance structure have boosted the development of GE interaction models, which are described in details later.

The same trait measured in different environments can be treated as several correlated traits (Falconer and Mackay 1996). For example, grain yield measured in one environment and in second environment can be treated as two different traits. If these traits are highly correlated, there is little evidence of GE interaction, while if correlation is low, GE interaction has supposedly highly significant effect in the population. Robertson (1959) suggested that appropriate limit for correlation between two environments containing GE interaction is below 0.8.

1.3 GENOMICS IN PLANT BREEDING

Estimation of breeding values is one of the primary interest of breeders. Best linear unbiased prediction (BLUP) was introduced in the context of animal breeding (Henderson 1963, Goldberger 1962, Henderson 1975) in order to improve the selection index approach. For genetic random effects, a variance-covariance matrix is included to account for relationships between animals and it allows information between relatives to be shared. For plant breeding, BLUP was introduced only during the 1990s (Bernardo 1994, Panter and Allen 1995a, 1995b) and did not gain such a popularity as in animal breeding. Reasons were speculated in Piepho et al. (2008) as follows: firstly, in plant breeding phenotypic records per genotype are mostly adequate, because of the repeated measurements of the lines in multiple years and environments. Therefore, BLUPs or best linear unbiased estimates (BLUEs) seemed not to differ substantially. Second, BLUPs in animal breeding are sometimes a must, due to lack of direct measurement of a trait, such as milk production in sires, whereas in plant breeding, there are hardly any similar cases, except combining ability. Third, in animal breeding, the number of genotypes is often large and, because of that, genetic variance can be accurately estimated, whereas this is generally not the case in plant breeding, where the number of genotypes is limited. In self-pollinated crops, pedigree-based BLUPs arising from historical pedigree records have been used for soybean. BLUPs were found to be more accurate than standard method, in this case a mid-parent

value (Panter and Allen 1995a). Similar studies have been conducted for barley (Bauer et al. 2006) and wheat (Oakey et al. 2006), where pedigree-based BLUPs were compared with BLUPs without the relationship information. Results showed that selection of parents based on pedigree-BLUPs was more efficient.

The approaches of BLUP and pedigree-BLUP have been enhanced by using genetic marker data to describe the relationships between individuals (VanRaden 2008). This approach is known as genomic BLUP (GBLUP) and results in the prediction of genomic estimated breeding values (GEBV). The approach is designed for genomic selection (GS), where marker effects within a population are estimated and used to predict GEBVs of untested population. The aim is not to find specified alleles affecting the trait, but rather accumulating positive alleles through summing marker effects (Meuwissen et al. 2001). GS has been found to be a promising method for dealing with quantitative traits (Bernardo 2016). In contrast, genome-wide association study (GWAS) is performed in order to search for specific alleles within the collection of individuals by conducting inference for each estimated marker effect. Significant associations between traits and markers can be found with limited number of quantitative trait loci (QTL). However, detection of loci with small effects or rare alleles is not feasible with GWAS. Both methods, taking advantage of marker information, are discussed in details in the following chapters.

1.3.1 GENETIC INFORMATION

In molecular biology, it is possible to quantify genetic variation in the DNA level and applications using marker-based methods have become common (Lynch and Walsh 1998). Candidate loci of the biological process is often unknown, but QTL can be detected indirectly through linked marker loci. In the book by Lynch and Walsh (1998) it was mentioned that it is routine to have 50 to 200 molecular markers for any species of interest. Twenty years later, the number of markers available for analysis has reached thousands, even million for some crops like maize. This has been the result of progress in molecular technology and considerable decrease in costs of sequencing and genotyping. Today, covering the whole genome with markers and even sequencing of the whole genomes is possible.

Single nucleotide polymorphism (SNP) markers are currently the most used molecular-marker system. The success of SNPs has been in adaptation of high-throughput technologies, allowing a large number of DNA samples with large number of markers to be processed efficiently (Rafalski 2002, Hyten et al. 2008). Price of SNP genotyping technology continues to drop as genotyping-by-sequencing (GBS) is used (Poland and Rife 2012). More information on chromosome positions of the SNP markers is becoming available and eventually whole-genome sequences for crops are been

published, as have been for barley (Mayer et al. 2012, Mascher et al. 2017) and wheat (IWGSC 2018).

1.3.2 GENOMIC SELECTION

GS was first introduced to animal breeding (Meuwissen et al. 2001, Hayes et al. 2009), and from there it gradually spread to plant breeding research (Bernardo and Yu 2007, de los Campos et al. 2009, Crossa et al. 2010, Crossa et al. 2011, Massman et al. 2013) and to commercial plant breeding programs (Nielsen et al. 2016, Kristensen et al. 2019, Michel et al. 2016). The core of GS is a training population, which contains a set of individuals with phenotype and genotype information. A prediction model is fitted for individuals in a testing set with only genotype information to receive GEBVs. Simulations have shown that GEBVs were as accurate as traditional EBVs generated from progeny testing in cattle (Hayes et al. 2009). Progeny testing takes years to implement due to the prolonged generation interval in animals. In breeding, prediction of GEBVs would mean doubling the rate of genetic gain and 92% savings in the process when progeny testing is not needed anymore (Schaeffer 2006).

1.3.2.1 *Methods of genomic prediction*

Prediction of GEBVs can be done with multiple models. One of the most common approaches is GBLUP, which originates from pedigree-BLUP, the mixed model approach based on pedigree-derived relationships in animal breeding (Henderson 1975). In GBLUP, the pedigree-derived relationships can be replaced with a relationship matrix calculated from genetic markers (VanRaden 2008). The base form of the model equation can be written as:

$$(2) \quad \mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

where \mathbf{y} is the vector of observed phenotypes. Vector \mathbf{b} corresponds to fixed effects, vector \mathbf{u} to random effects (GEBVs) and vector \mathbf{e} holds the random residuals. The design matrices \mathbf{X} and \mathbf{Z} associate the observations to the fixed and random effects, respectively. Common assumptions for the mixed model are: \mathbf{u} follows a multivariate normal distribution $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G}\sigma_a^2)$ where \mathbf{G} is the marker-derived relationship matrix and σ_a^2 the additive genetic variance, $\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$ where \mathbf{I} is the identity matrix and σ_e^2 the residual variance and covariance between (\mathbf{u}, \mathbf{e}) is assumed to be zero.

An equivalent genomic prediction model can be obtained by predicting single marker effects instead of GEBVs, when \mathbf{Z} associates genotypes to a vector \mathbf{u} of holding all marker effects. This alternative formulation of the prediction model was introduced as ridge regression BLUP (rrBLUP) (Whittaker et al. 2000). With a large number of markers, prediction models

like rrBLUP explicitly aiming at predicting single marker effects face the so-called “large p – small n ” problem, where the number of the predictors is much larger than the number of observations. All predictors cannot be estimated simultaneously due to a lack of degrees of freedom. Another problem is that markers are correlated and impose a multicollinearity problem leading towards overestimation of marker effects (Lorenz et al. 2011). These problems can be treated by using so-called shrinkage models: approaches are setting shrinkage factors with random regression (Meuwissen et al. 2001), variable selection models referred to as least absolute shrinkage and selection operator (LASSO) (Tibshirani 1996), kernel methods like reproducing kernel Hilbert spaces (RKHS) (Gianola et al. 2006), where the information of predictor variables are converted into a matrix with fewer dimensions, corresponding to marker data converted into a relationship matrix, and dimension reduction methods, like partial least squares (PLS) (Wold 1985) and principal components (PC). A Bayesian framework can be used to lessen the stringent assumption of equal marker effect variances in rrBLUP (Meuwissen et al. 2001) while random regression assumes equal variance for all predictors, in the group of models referred to as the Bayesian alphabet (Gianola et al. 2009), such as A, B (Meuwissen et al. 2001), C π (Habier et al. 2011), R (Erbe et al. 2012), variable distributions for predictors are assumed. Different prediction methods have been compared in multiple studies (Heslot et al. 2012, Maltecca et al. 2012, de los Campos et al. 2013). In general, when a trait with a polygenic architecture is predicted, the differences between models are small, but when a few QTL control a trait (Anderson et al. 2001, Munkvold et al. 2009), then using Bayesian methods should be considered (Lorenz et al. 2011). Using kernel-models, describing non-additive effects, has been shown to be beneficial in multi-environment predictions and in the presence of GE interaction (Cuevas et al. 2018, Sousa et al. 2017).

The GBLUP model can be extended to be used for multi-trait prediction. The multi-trait prediction can take advantage of correlated traits with higher heritability than in the trait of interest. This principle originates from indirect selection (Falconer and Mackay 1996). The information is shared between the traits and prediction has been shown to improve (Jia and Jannink 2012, Rutkoski et al. 2016). On the other hand, there are studies that show no substantial increase in prediction ability (Kristensen et al. 2019, Schulthess et al. 2016).

GE interaction complicates genomic prediction within plant populations and has shown to lower predictability (Ly et al. 2013, Dawson et al. 2013). With recent development in statistical methods, GE interaction can be incorporated into prediction models with the aim of using the information shared between environments and, as well, deepen understanding of underlying GE interaction patterns (Jarquín et al. 2014, Burgueño et al. 2012, Heslot et al. 2014). With high-dimensional genotypic and environmental data, it quickly becomes unfeasible to assess the interaction term. In order to manage the high-dimensionality Jarquín et al. (2014) suggested a reaction norm model

framework, where a variance-covariance structure generated from genetic markers and environmental covariates is used to model the GE interaction term. Heslot et al. (2014) used a growth stage synchronized stress covariates to learn more on architecture of GE interaction and underlying stress environments. They were able show the most influential environmental stages causing stress for their testing sites.

1.3.2.2 Accuracy of genomic prediction

Accuracy of genomic prediction is defined as the correlation between GEBVs and corresponding true breeding values (TBVs). As TBVs are not known in real data sets, accuracy must be approximated by using surrogate estimates for the TBVs: e.g. pseudo-phenotypes; i.e. summary estimates for the observed phenotypes calculated across trials, or within trial observed values. Pseudo-phenotypes can be constructed by adjusting observed values for trial-specific effects and averaging these adjusted values. When comparison between predicted and observed values is made within each trial, focus is in understanding GE interaction of tested lines in different environments. When comparison is made between GEBVs and pseudo-phenotypes, interest is more in prediction success across environments.

Validations are done with resampling methods, where samples are drawn from a data set and repeated, in order to compute the direct correlation between predicted and observed values or mean-square error (MSE) (Verbyla et al. 2009). The most common resampling method is cross-validation (CV) (Lorenz et al. 2011). In the simplest case, a data set is divided randomly into two equal sized sets, a training and a test set. In leave-one-out cross-validation, one observation of the n observations is left to test set and others are used in a training set. Validation process is repeated for n times. Observations can be divided into k -folds, where one of the folds is used as a test set and others in a training set. Validation procedure is repeated at least k times (James et al. 2013). Validations have been executed in more stratified selection of dividing into a training and a validation based on families, breeding cohorts or breeding cycles (Kristensen et al. 2018, Michel et al. 2016). The appropriate validation method should be chosen to meet the need, which can be previously mentioned prediction ability across families, cohorts or breeding cycles or within breeding cycle ability e.g. ability to predict full sibs for another environment. Burgueño et al. (2012) validated prediction with two schemes, by computing accuracy, for lines which were not evaluated in any observed environments (CV1), and for lines which had some observations from predicted environments (CV2). Jarquín et al. (2017) added CVo for untested environments and CVoo for untested environments and untested lines.

Forward validation can be computed by chronologically dividing the data set into a training and a testing set (Bernal-Vasquez et al. 2017, VanRaden et al. 2009). Resampling can be used within the training set, but predefined set can be used as a single measure of predictability. Forward validation describes

the next generation or breeding cycle and year. It is used to validate the across-cycle selection accuracy.

1.3.2.3 Factors affecting prediction accuracy

As indicated above, the prediction model chosen, GE interaction and the validation method of estimating the level of accuracy affects the estimate of accuracy. Additionally, the trait heritability together with the size of the training population are key factors that influence prediction accuracy. In the early stage of implementing GS, genotyping is most likely a considerable investment and, for economic reasons, it might be necessary to limit the size of the initial training population. Accuracy has widely been shown to increase with larger training population (Nielsen et al. 2016, Meuwissen 2009), but for traits with simpler inheritance, similar level of accuracy can be reached with smaller training population (Hayes et al. 2009, Lorenz et al. 2011). Besides heritability, optimal size depends on multiple factors, such as crop species, breeding program and breeding aim. Suggested training population sizes are around hundreds to thousands: 700 (Cericola et al. 2017) or 2000 (Norman et al. 2018) wheat breeding lines with varying relatedness. Bassi et al. (2016) suggested that, if the desired level of accuracy is above 0.5, then the training population should contain 50 full-sibs, 100 half-sibs or at least 1000 individuals with a more diverse background in relation to the breeding population.

A training set should not only consist of close relatives, but should represent the whole breeding population (Isidro et al. 2015). Relatedness between the training and test set increases accuracy more than size of the training population (Edwards et al. 2019). As a result, if the size of the training population is fixed, it should contain more crosses with less siblings rather than few crosses with more siblings. It has been shown that more distant individuals, breeding cohorts or cycles decrease accuracy (Lorenz and Smith 2015, Nielsen et al. 2016), because the accumulation of recombination events over generations corrodes the linkage pattern between markers and QTL present in the training population. With random mating the rate of accuracy loss per generation is 5% (Meuwissen et al. 2001), but more if selection is involved (Muir 2007). Therefore, the prediction model should be updated with each breeding cycle (Heffner et al. 2010), while Jannink et al. (2010) suggested that at least the parents of each breeding cycle should be phenotyped, because the next generation carries only the alleles of their parents.

Comparisons between pedigree-based BLUP and marker-based GBLUP have been made in CIMMYT with wheat breeding data sets. Some results show that at least for prediction ability the results differ negligibly (Juliana et al. 2017) for some traits and more (7.7-35.7%) for others (Crossa et al. 2010). Among the reasons listed in Juliana et al. (2017), were: if pedigree is deep, as in the case of the CIMMYT wheat breeding program, and family size is small, the benefit of using markers instead of pedigree records is not substantial.

When using pedigree information only, the Mendelian sampling term, which gives rise to differences between full sibs, is overseen (Daetwyler et al. 2007, Hayes et al. 2009, Crossa et al. 2011) and actually prohibits selection based on pure predictions within crosses.

1.3.2.4 Applications of GS

The studies on GS mostly involve technical aspects of prediction, but there are some studies on how the implemented GS has affected the breeding process (Bernardo 2016, Massman et al. 2013, Asoro et al. 2013, Rutkoski et al. 2015, Beyene et al. 2015, Combs and Bernardo 2013). Comparisons have mainly been made with MAS (Massman et al. 2013, Asoro et al. 2013) and BLUP-based phenotypic selection (Asoro et al. 2013, Beyene et al. 2015, Combs and Bernardo 2013). Both types of comparisons show that GS is the more preferable method of selection. Rutkoski et al. (2015) pointed out that even though GS reached similar genetic gain as phenotypic selection, the amount of genetic variance was significantly decreased in GS compared to phenotypic selection, which is a considerable problem in a long-term breeding program.

1.3.3 GENOME-WIDE ASSOCIATION STUDY

GWAS uses genome-wide markers to estimate marker effects and test marker significance on phenotype. When comparing GWAS to traditional QTL mapping approach, several differences can be found. In GWAS, a broader population is used instead of only biparental progeny, and therefore generalization of association results outside the studied population has shown more promise. Specific mapping populations, which demand time, costs and effort to create, are no longer compulsory (Bernardo 2016). GWAS originates from human genetics (Hästbacka et al. 1992, Risch and Merikangas 1996, Altshuler et al. 2008, Donnelly 2008) and has been gradually adopted by plant research (Thornsberry et al. 2001, Nordborg and Weigel 2008). The causal alleles are hardly ever found in GWAS, but the key is that at least one of the genetic markers is in linkage disequilibrium (LD) with causality (Myles et al. 2009). As realized result of GWAS, the associated marker can be used as a starting point for revealing the causal mutation by comparative genomics, fine mapping or other genomic approaches. The associated marker can be used as a selection marker in the breeding process even without uncovering the causality. Subsequently, the use of markers aims at enriching favorable alleles in the breeding population. The durability of LD in multiple generations defines the usability of a selection marker as LD between the trait and the associated marker can be broken by recombination.

Benefit of GWAS in public plant breeding is controversial as Bernardo (2016) stated that new candidate gene discovery has been limited with GWAS compared to linkage analysis and none have been introgressed into elite

germplasm. He speculates that common variants are discovered, but GWAS fails to identify rare variants, which would be of interest, especially in the case of disease resistance. When the aim is to identify common variants existing in the population, GWAS is more promising. When marker discovery within a breeding program is executed in order to find superior alleles and to enrich them via selection, GWAS is a valid approach. (Zhu et al. 2008). In human genetic studies, benefit of GWAS has been reviewed extensively (Naidoo et al. 2011, Donnelly 2008).

1.3.3.1 Methods of GWAS

In simplified form, GWAS can be formulized as N number of “independent” hypothesis tests, where N is the number of markers. Tests are not truly independent as markers can be correlated. When the null hypothesis is rejected, it can be assumed that a marker correlated with a causal polymorphism has been found. As GWAS tests multiple times, in order not to increase type 1 error, multiple testing correction is used. Common methods are Bonferroni testing correction and false discovery rate (FDR) (Benjamini and Hochberg 1995). Association analysis can be executed with multiple methods. The simplest one is to test the significance of the marker effect using markers as fixed effect in a naive model. Population structure can be treated by adding covariates, which describe the structure, as fixed effects into the model (Pritchard et al. 2000). This model is commonly called Q model (Yu et al. 2006, Arruda et al. 2016). Yu et al. (2006) described QK models, where mixed models are used to test if marker effects can be detected with threshold p-value. Markers and population structure are defined as fixed effects and kinship (**K**) matrix calculated from markers is treated as a random effect, accounting for relatedness.

Bayesian framework has been used in association studies (Pikkuhookana and Sillanpää 2014, Kristensen et al. 2018, 2019, Marttinen and Corander, 2010). The advantage of Bayesian inference lies in unequal variance assumption for predictors. In infinitesimal models, such as rrBULP, all of the predictors are shrunk with similar intensity, but Bayesian approach allows different shrinkage for predictors. This might be useful for traits, which can be largely be explained by a few predictors and most of the predictors have zero effect.

1.3.3.2 Factors affecting GWAS

The power to detect association depends on the size of the studied population, allele effect size, density of markers, rate of LD decay and the decided level of significance (Gordon and Finch 2005). Simulation studies show that the size of the population and repeated measurements increase power to detect QTL (Arbelbide et al. 2006, Yu et al. 2005). Another simulation study showed that

the number of phenotyped lines increases power more than the density of markers (Long and Langley 1999). The same study suggested a sample size of 500 individuals to be appropriate for analysis, while many studies conducting GWAS have minimal sample size of 100 (Zhu et al. 2008). The minimal sample size is hard to set due to other dependencies of the analysis, but through simulations it was shown that large sample size is required to obtain high power for detecting moderate allele effects (Zhu et al. 2008).

Relatedness and population structure within the studied populations may lead to spurious associations (Lander and Schork 1994) and high false positive rates (Aranzana et al. 2005). This is caused by phenotypic variation between populations being highly correlated with difference in allele frequency in these populations. Relatedness and population structure can be treated separately. Within a breeding population relatedness is commonly high and to account for it, a relationship matrix has been used (Yu et al. 2006). The two most common ways to correct for population structure are by STRUCTURE (Pritchard et al. 2000), where the studied population is divided into hypothetical subpopulations, and by principal component analysis (PCA) (Price et al. 2006, Patterson et al. 2006), where dimensionality is reduced and eigenvectors for describing the most variation within the studied population are computed. From these, PCA has the advantage of being assumption-free (Price et al. 2010). Zhao et al. (2007) found that both methods were able to capture underlying structure reasonably well.

GWAS is largely based on correlation between the causal allele and an associated marker allele. Strong correlation indicates close linkage and LD is stronger. Recombination breaks down LD, and LD decay, which is the return of an association to random between two loci by time, has been used to quantify the marker coverage needed to perform GWAS. LD decay varies extensively between crop species and is found to be low in self-pollinated species (Malysheva-Otto et al. 2006, Remington et al. 2001, Myles et al. 2009), especially in breeding populations (Bengtsson et al. 2017) where selection can affect formation of large LD blocks. LD decay varies between different chromosome regions, as recombination rate is higher further apart from the centromere of the chromosome (Flint-Garcia et al. 2003).

As discussed before, rare variants are problematic for GWAS. Based on population genetic studies, a major portion of all polymorphisms is due to rare variants (Gibson 2012), with frequency less than 0.5% (Hartl and Clark 2007). Therefore, it is problematic to cover the phenotypic variation, which as a part leads to the concept of “missing heritability” (Manolio et al. 2009). Most of these rare alleles will more likely be undetected, because of low statistical power. If rare alleles are known to carry potentially usable variation, designated crosses, to increase allele frequency in the population, should be considered to provide power to detect variation. Such an approach was used in maize when nested association mapping (NAM) population was developed for detecting marker associations (Yu et al. 2007, Myles et al. 2009). Multi-parent advanced generation inter-cross (MAGIC) populations have also been

used to enrich favorable alleles for GWAS (Sannemann et al. 2015, Cavanagh et al. 2008).

2 AIMS OF THE STUDY

The main aim of this study was to investigate usability, robustness and practical operations of genomic selection within different breeding dilemmas. To reach this aim, example breeding data sets from oat and barley were used in the analysis. Genetics behind the data sets were uncovered with SNP markers, and genomic prediction, genome-wide association study and study on GE interactions were conducted. The study was concentrated on ‘difficult’ traits in plant breeding to increase knowledge surrounding these topics.

The specific aims of this study were:

- 1) to improve prediction of grain yield with (trait-assisted) multi-trait prediction compared to single-trait prediction for oat and barley. Genetic correlation between grain yield and maturity/protein content were expected to give higher accuracy for grain yield in multi-trait prediction (publication I).
- 2) to study genetic effects of Fusarium head blight resistance related traits in oat with genome-wide association study and conduct genomic prediction for the traits. The primary hypothesis was that unlike wheat, oat does not have major genes for resistance within the breeding material and genomic prediction would present a more practical approach to assist resistance breeding (publication II).
- 3) to study and compare the effect of GE interaction within the breeding programs of oat and barley. GE interaction was included into the prediction model in order to test if prediction ability of grain yield could be improved. The prediction ability was expected to increase. The second objective was to study within year correlation between trial locations in order to structuralize the testing network. The hypothesis was that GE interaction exists to some level, but no negative correlations would be found (publication III).

3 MATERIALS AND METHODS

Details from materials used in each original publication are listed in Table 1. More detailed description of the materials used in the analysis are given in the original publications I-III.

Table 1 *The data set details used in each original publication I-III.*

Publication	n Lines¹	n Families²	n Markers³	n Obs⁴	n Envs⁵	n Years⁶
I: oat	2 043	491	3 905	9 610	401	10
I: barley	4 262	523	5 813	20 737	218	10
II: oat	327	162	2 785	797	7	4
III: oat	1 894	397	3 786	4 573	40	8
III: barley	3 997	452	5 813	10 434	65	9

¹ the number of plant lines; ² the number of unique families from which the lines were derived from; ³ the number of markers after filtering; ⁴ the number of observations of the trait studied. The number of observations for publication I is grain yield, publication II DON and publication III grain yield; ⁵ the number of location*year combinations; ⁶ the number of years presented in the study.

3.1 BREEDING OF SELF-POLLINATED CROPS

The data sets for spring oat and spring six-row barley used in the original publications were generated within a breeding process. In both species the breeding populations were generated by bi-parental crosses. For barley, haploid microspores of produced F₁ progeny were used in DH tissue culture to produce fully inbred breeding lines. For oat, inbreds were produced through multiple generations of SSD. It should be noted that it takes several generations before seed has been propagated enough and sowing of multi-environment trials is possible as the seed multiplication of a breeding line starts from one DH plantlet for barley, or one seed for oat. The first year of field testing was executed as head rows for both crops (oat F_{4:5} and barley DH₁), where only agronomic traits were selected. The first year of grain yield testing was done in preliminary yield trials (PYTs). At this stage, the seed maybe sufficient for one replication of grain yield plot at one location or seeds were in some cases multiplied in shuttle breeding using contra-season nursery, and multiple locations of PYTs were sown. Selection for grain yield started from these multi-location trials for both crops (oat F_{4:7} and barley DH₃). During following years, advanced yield trials (AYTs) in multiple locations were sown and selection continued to reduce the amount of tested cultivar candidates.

3.2 BREEDING DATA SETS

The oat and barley breeding data sets (publication I) contained observations of grain yield (kg/ha), maturity (a visual score from 1 to 5 for oat, where 1=early and 5=late, for barley, the scores were from 1 to 7) and protein content (%) for 2043 oat and 4262 six-row barley breeding lines and check cultivars. The data were recorded between 2008 and 2017 and the lines represented seven oat and nine barley breeding cohorts each consisting of multiple bi-parental populations.

The oat data set (publication II) for FHB resistance related traits (DON= deoxynivalenol (ppm), qFUSG= *F. graminearum* pg DNA (relative to oat 100 ng DNA) content, FIK= *Fusarium* infected kernels (%) and GC= germination capacity (%)) and agronomic traits (HD= heading date (days), PH= height (cm), MC= maturity (1-5)) was collected from artificially spawn inoculated field experiments. The data set contained 327 breeding lines, accessions from gene bank and check cultivars recorded between years 2015 and 2018.

The oat and barley breeding data sets (publication III) contained observations of grain yield (kg/ha) from PYTs for breeding lines and check cultivars. The data sets were collected between 2010 and 2018. The weather data were gathered from the national weather observation stations (Finnish Meteorological Institute) closest to the trial locations. Environmental covariates for weather variables were calculated by dividing the growing season into 10-days intervals starting from the sowing date of each trial. Soil type, growing zone information and computed weather covariates were merged into a data table, which was used to calculate a relationship matrix for the trials.

3.3 GENOTYPES AND POPULATION STRUCTURE

Genotypic information for the original publications I-III was generated with custom SNP marker chips. Quality of markers was checked to meet a call rate of 90%, a minor allele frequency (MAF) of more than 1% (I, III) or 5% (II), and heterozygosity of less than 5%. The genomic relationship matrix (**G**) was calculated (I-II) using the first (I, II, III) and the third (III) equation from VanRaden (2008). For genomic prediction (I-II), an inverse of the **G** matrix was taken prior to analysis (MiX99 Development Team, 2015). For GWAS, the relationship matrix to account for close relatedness was calculated using kinship (**K**) matrix arising from genetic marker information.

The population structure of both oat and barley breeding populations was computed in I-II with PCA. In oat, there was no clear separation into groups, but early and late lines showed some diversification on PC1. In barley, the differences between early and late lines were not as clearly observable on PC1. Due to presumable population structure, three first PCs were included in oat GWAS (II) to account for it.

3.4 VARIANCE COMPONENT ESTIMATION AND GENOMIC PREDICTION

For original publications I-III variance components were estimated using Average Information REstricted Maximum Likelihood (AI-REML) (Jensen et al. 1997). The estimated components were used to calculate narrow sense heritability for each trait. Trait correlations were assessed by applying multi-trait models with two (II, III) or three (I) traits. For the III, variance components were as well generated with six different models conducting effects arising from environment, line, genetic, interaction between genotypes and environments, environmental covariates and interaction between genotypes and environmental covariates.

In I, the estimated variance components were used to solve mixed model equations of single and (trait-assisted) multi-trait models to obtain GEBVs for unphenotyped lines. In the case of trait-assisted multi-trait prediction, the predicted lines had phenotypes from correlated traits; i.e. maturity and protein content, available for prediction. In II, the estimated variance components were used to predict GEBVs with single-trait mixed model. The III compared variance components estimated with the six different models.

3.5 VALIDATION OF PREDICTION

Genomic predictions were validated with various schemes. In I, fivefold cross-validation based on random division and stratified family division were used. Additionally, forward prediction where the data set was chronologically split into the training and testing sets was computed and repeated with cumulative data for each breeding cohort. In II, random fivefold and leave-one-out cross-validation schemes were used. In leave-one-out, each trial was set as a testing set one at a time. In III, random fivefold cross-validation was used in two alternative schemes, CV1 and CV2. The scheme CV1 presents the classical prediction aim, when the predicted line has no own phenotypic observations. In CV2, the predicted lines have some phenotypic information from other testing locations. The aim of this validation, is to test if some observations can be dropped without impairing too much of the prediction accuracy.

3.6 ASSOCIATION STUDY

In original publication II, GWAS was conducted for FHB resistance related traits as well as for collected agronomic traits. In GWAS, four different models were used: naive model, where no population structure or relatedness between lines were accounted for; Q model, where only population structure with three PCs was used; K model, where only relatedness between lines was corrected with a kinship matrix; and QK model, where both population structure and relatedness were accounted for. The results were examined through

Manhattan and qq-plots. To correct for multiple testing problem, both Bonferroni and FDR were used.

4 RESULTS AND DISCUSSION

The data sets used in the original publications represented oat and barley breeding programs. The data sets generated in a breeding program are large in size but not optimally constructed for research purposes. The breeding populations are exposed to selection, which makes breeding data biased in the sense, that better cultivar candidates are phenotyped many more times than low performing candidates. Due to selection, data sets are highly unbalanced as majority of the candidates are discarded after their first observations. In addition, new candidates enter the testing network every year. The situation is challenging as genetic variation in the breeding cohort diminishes through testing years and minority of the breeding lines receive a reliable estimate for a trait based on multi-year trial data. This makes the validation of the predictions demanding as the estimate of the TBV may be unreliable.

The data sets used in the original publications were not specifically generated for this thesis or limited to a certain breeding cohort. Therefore, the data sets may be weaker in the sense of research, as the data sets were unbalanced and limited mostly to breeding lines, not a broad collection of lines from a diversity panel. However, the strength of the analysis in the original publications is that the results are close to practice and can directly benefit the breeding process. In addition, the amount of data used in publications I and III could not have been generated for research purpose only. The methods useful for breeding are discussed in the following chapters in the context of each publication and, then the overall benefit of genomic approaches in breeding is summarized.

4.1 IMPROVING PREDICTION OF GRAIN YIELD

In the original publication I, genomic prediction of grain yield was studied and validated. The single-trait model showed a prominent level of cross-validated prediction accuracy for both crops. The accuracy was not significantly increased when multi-trait model was used. However, in the forward validation an increase in accuracy was seen when a multi-trait model was used instead of a single-trait model. The difference was not substantial or significant, but rather consistent in direction. In trait-assisted models grain yield prediction was assisted with correlated traits: maturity, protein content or both. The difference between single-trait and trait-assisted models was more apparent compared to previous comparison. However, trait-assisted models differed in consistency when compared with multi-trait prediction without assisting: the best prediction model varied between years. On average, if one trait was used to assist multi-trait prediction, grain yield accuracy was increased by 10% for oat with maturity and by 26% for barley with protein

content. These results were in-line with previous research (Fernandes et al. 2018, Kristensen et al. 2019). Even though, the increase in accuracy was less in our study compared to theirs. In Fernandes et. al. (2018) increase in accuracy was 50% and in Kristensen et al. (2019) 73%. The difference in the level of response could arise from differences in the heritability of the investigated trait. In Kristensen et al. (2019) trait heritability was higher (0.51 for protein content), than single-trait heritability of 0.33-0.39 for grain yield in our study. In comparison, in Fernandes et al. (2018) the single-trait heritability was lower (0.26 for biomass). Therefore, only the heritability does not explain our lower success. If grain yield prediction accuracy can be increased via a correlated trait that is considerable cheap to measure (e.g. maturity), the benefit is substantial, even if the increase in accuracy is modest. In the original publication, we were able to show that prediction of grain yield can be improved by (trait-assisted) multi-trait models. In summary, the benefit is in more accurate GEBV estimates for grain yield used in selection. In the original publication, we also showed that a training population can be constructed from a historical breeding data. Even without additional field tests for a specifically designed training population, we were able to compute reasonable prediction accuracies for grain yield.

One of the key results from the forward validation, was that breeding cohorts showed considerably varying levels of accuracy for grain yield. The variation of the cohorts can be due to the use of genetically distant market cultivars as crossing parents. This brings new genetic variation to the population, but at the same time can cause lower genetic relationship between breeding cohorts (Michel et al. 2016, Cericola et al. 2015). More likely, the predominant cause of variation was the environment. The testing location remained rather constant, but variation in yearly weather had a large effect on field trials. This result emphasizes the need to more accurately define the causes and effects of different testing years. The yearly difference also shows that multiple breeding cohorts should be used as a training population.

Grain yield prediction is demanding but still the most important goal in a commercial breeding program. In the publication we were able show that accuracy of prediction can be increased by using trait-assisted or multi-trait models. Most of all, including correlated traits into the prediction model did not decrease accuracy in any relevant case for grain yield. In the early generations, grain yield is not selected per se (head rows) or selection is inaccurate due to limited number of observations. In this situation, observations on maturity and measurements of protein content could potentially increase prediction accuracy for grain yield, when the predicted lines receive their GEBVs. Using trait-assisted GEBVs may able selection for grain yield to be more accurate from early generations.

4.2 GENOMICS OF FHB RELATED RESISTANCE TRAITS

The original publication II was predominantly performed to learn more on a challenging breeding objective, like FHB resistance in oat. FHB resistance is a complex feature and includes many traits which together enhance the resistance against the disease and toxin accumulation. Resistance has been shown to correlate with agronomic traits, such as height and maturity in wheat (Buerstmayr et al. 2009, Emrich et al. 2008, Lu et al. 2013). The first objective of the study was to report genetic correlations in Finnish conditions for the agronomic and resistance related traits. The results showed high genetic correlations between FHB resistance related traits (DON, qFUSG, germination capacity, *Fusarium* infected kernels), but poor correlations between FHB resistance related and collected agronomic traits (plant height, heading date, maturity class). Therefore, we were not able to confirm, that the well known pattern that taller and late cultivars have less FHB in wheat, also would apply in oat. All the FHB resistance related traits had low estimates of heritability, which may imply that the amount of observations to estimate variance components reliably was too limited. In our study, we might have lacked an extremely susceptible check, which potentially affected the genetic variation estimated. In Bjørnstad et al. (2015), their material included unadapted Midwestern cultivars, which showed an extremely susceptible phenotype. They speculated that it might have been due to photosensitive response. In their study, all Nordic cultivars were placed in the middle of resistance response to FHB and DON accumulation. FHB resistance might therefore be partly affected by photosensitive response. FHB resistance in oat is generally highest in Nordic breeding material or at least in the same level with known moderate resistance (MR) sources (Hautsalo et al., unp. results), like Leggett and Stride (Mitchell Fetch et al. 2013).

In GWAS, we were not able to detect, with reasonable control of population structure or relatedness, any significant marker-FHB resistance related trait associations. We were able to show significant associations for markers and maturity with K model, and a peak, which was clear but not significant, for *Fusarium* infected kernels in the studied data set. The low success in associations indicates that the genetic basis of *Fusarium* and DON resistance may be polygenic, based on a high number of loci with small effects. The more preferable method for using genome-wide information in FHB resistance breeding could thus be genomic selection. In another words, while there were no potential associated marker selection candidates, all markers to describe the differences between lines, should be used. Use of genome-wide marker information through GS is, therefore, currently the only potential option to benefit from genetic information. Computed GEBVs can be used directly for selection, but model validation should be emphasized when deciding on selection method. The results from the model validations for FHB related traits showed rather low accuracies. From these traits, germination capacity

showed the highest accuracy and computed GEBVs are usable in selection. For the other traits, the GEBVs generated should be used with caution, or used only when discarding extremely susceptible cultivar candidates. Closer comparison between the candidates is questionable, since small differences in GEBVs are not reliable. More data should be generated in order to improve the prediction and reliability of GEBVs. In future, the genetic correlations between FHB resistance traits should be used for computing GEBVs based on multi-trait prediction models to see if prediction of the resistance traits can be improved.

4.3 GE INTERACTION WITHIN BREEDING POPULATIONS

The original publication III studied the quantity of GE interaction within tested trial locations in PYT for multiple breeding cohorts. The study was performed in two stages. The first stage presented additive genetic correlations between testing locations within each year for a breeding cohort. After computing these yearly correlations, a mean across years was taken from correlations between testing locations repeated during multiple years. A mean per trial location pair was received, but the number of correlations varied between one and 18 depending on how many times two trial locations occurred within a same year for a breeding cohort. In the second stage, genomic prediction was done for all data and prediction accuracy was quantified with two cross-validation schemes (CV1 and CV2). Prediction was conducted with six different models as detailed in the original publication.

The results of the yearly additive genetic correlations showed that in oat the amount of the data was limiting the results. Only five location pairs were repeated across years more than two times. For the comparisons with three records results were above 0.9, which may be due to too limited comparisons or no GE interaction occurring. For the two other comparisons, the trial locations may have experienced specific stress circumstances and the correlations were therefore low. For barley, more reliable results were received. The mean correlations varied between 0.41 and 0.77 for the records which had more than seven comparisons behind them. These results imply that GE interaction is present, but the extent is not so large that it would produce negative mean correlations. Negative correlations were occasionally observed for single years, but the mean values were not close to zero.

The second stage included an analysis of computed environmental covariates to describe the similarity between the trial locations. The drawn dendrograms showed that while oat had less trial locations within a year, the dendrogram clusters captured year effects more clearly than for barley, which had more locations per year. The variance components computed with the six models revealed differences in the effects included in the models. The most relevant research findings were that barley had more than twice as large

variance due to genetic effect than oat, and barley also had larger GE interaction effect. The variance due to environmental covariates was insignificant, but the interaction between genotype and environmental covariates was larger. However, including interaction between genotypes and environmental covariates in the prediction model was not beneficial compared to models where interaction between genotype and environments was treated as unrelated. The prediction accuracy was lower and the residual effect was higher in the EGW-GxW model compared to the model with interaction due to the genotypes and environments (EG-GxE model). Overall, the residual effect was lower when more parameters were included in the model, especially when interaction between genotypes and environments was used. Additionally, the accuracies of the models containing interaction between genotype and environments (EG-GxE) were the highest for both crops. The increase in accuracy was higher for barley, which is in line with the GE interaction variance components and the fact that barley had more locations within a year compared to oat. Overall, the accuracies received were on the same level as in previous studies (Burgueño et al. 2011, Pérez-Rodríguez et al. 2015, Jarquín et al. 2014) using the same models. In our case, CV1 accuracy was higher than in Pérez-Rodríguez et al. (2015) and Jarquín et al. (2014), but lower than in Burgueño et al. (2011). For CV2, the results were lower or on a similar level than in the cited research.

The results of this study show that in order to explore GE interaction and produce significantly relevant information for the breeding program, the data set should be larger than what we had for oat. The trial locations should be repeated in multiple years in order to take advantage of the correlations between locations, as it was observed that genetic correlations computed within a year may differ dramatically (SD of the mean across years). The true differences between locations are revealed after multiple years of simultaneous testing. For this reason, studying of the GE interaction within a breeding program is troublesome, but can be done if trial data is as extensive as in our barley data set. Repeated trial locations showed relevant correlations across years. These results show that GE interaction is present and should be taken into account while predicting line performances. Including GE interaction within the prediction model has shown to improve the prediction ability across tested environments (Burgueño et al. 2011, 2012, Pérez-Rodríguez et al. 2015, Jarquín et al. 2014), while accuracy of prediction is increased and residual variance is decreased.

4.4 APPLYING GENOMIC SELECTION TO BREEDING PROGRAMS

There are two main considerations when applying GS in breeding programs: when GS should be used and what is the most efficient way to use it in order to obtain the most practical improvement in the breeding process.

Optimization of the process should be done within specific conditions of each breeding program by considering the local growing conditions, the breeding material, practicalities of the process and, most of all, the available economic resources. In the literature, suggestions can be found, on how implementation of GS could be done (Gaynor et al. 2017, Hickey, et al. 2017, Crossa, et al. 2017, Bassi et al. 2016, Robertsen et al. 2019). Often when implementing GS, breeding process is divided into two components: across breeding cycle prediction i.e. population improvement part, and within breeding cycle prediction i.e. product development part.

In population improvement part, which was defined by Gaynor et al. (2017) and Hickey et al. (2017), potential parents to generate the next breeding cycle are identified and rapid recurrent GS is used to promote frequency of favorable alleles. The key aspect is in shortening of the breeding cycle in order to increase genetic gain (Longin et al. 2015, Schaeffer et al. 2006). A shorter breeding cycle has been described as one of the main advantages of GS (Crossa et al. 2017). In this context, a cultivar candidate, even from the first segregating generation, can be used as a parent as soon as it is genotyped and GEBVs have been calculated. The line may not have any phenotypes yet. In Gaynor et al. (2017), population improvement included crossing F_1 plants and doubling the number of crosses made yearly. Therefore, the breeding cycle time would be cut into half of a year. In their simulations, they were able to show the superiority of dividing the breeding process into two parts, namely population improvement part and product development part. The process produced 2.36-2.47 times more genetic gain compared to conventional breeding program with phenotypic selection and 1.31-1.46 times more gain over the conventional GS program. In the conventional GS program, the lines were genotyped at head row stage ($F_{4:5}$). The gain of two-part GS compared to conventional GS came from producing crosses twice a year. There are at least two caveats in the reduction of breeding cycle time to only 6 months. As in GS, also in such a rapid generation cycle, the risk of losing genetic variance is large. The initial loss of genetic variance is primarily due to more accurate selection of parents. Then the following loss is due to e.g. “effective population size, selection intensity, selection accuracy and the number of cycles per year” as listed in Gaynor et al. (2017). The second caveat is the loss of accuracy. When parents are selected many years before they are phenotyped (if they are phenotyped at all), this decreases the genetic relationship between the training population and the predicted lines, which leads to loss of accuracy. The prediction model should be updated with reasonable frequency, because the accuracy is decreased by 5% or more per generation (Meuwissen et al. 2001, Muir 2007). The suggestion from Jannink et al. (2010) that at least the parents of each breeding cycle should be phenotyped, should be studied if the extent is sufficient.

The product development part or within cycle prediction aim at identifying the superior crop cultivars as early as possible in the breeding process. GS has shown promising results (Schmidt et al. 2016, Michel et al. 2018) in improving

selection of traits, which are difficult to measure with low seed availability, such as grain yield and quality traits. In other words, early stage selection for many traits, especially the quantitative traits, is inaccurate in the breeding process and could potentially be improved by genomic selection. Grain yield or quality traits cannot be measured from head rows, and testing for these traits starts later in the breeding process. For grain yield the size of the plot is a limitation in head row stage. For quality traits the number of candidates is too large to be phenotyped with expensive measurements, e.g. beta-glucan content or FHB resistance related traits in oat. With genomic prediction, based on a training population that represents the genetic variation within the breeding program, GEBVs can be calculated for candidates as soon as they are genotyped.

Phenotyping costs in product development pipeline could be cut down with genomic prediction (Robertsen et al. 2019). Within a year, cultivar candidates are tested in multiple testing locations from PYT onwards. Based on computed cross-validation results (I), grain yield prediction accuracy will most likely not be affected by sparse testing, which means that not all lines are tested in all locations but instead a set of lines is tested in each location. Correlated traits which could easily be measured in the early phases of breeding could be used to rise accuracy of GEBVs for yield. This idea has been used in high-throughput phenotyping in order to improve prediction of grain yield (Krause et al. 2019).

Applying GS into the breeding program is a multidimensional problem, where specific components should be put together like a puzzle. Every crop behaves differently and has its own twists and need its own adjustments. The key to success is to find the most efficient approach and fit it for the particular breeding process. The breeding program budget is the frame, but everything else can be made accordingly. In Larkin et al. (2019), 14 different ways are detailed on how genomic selection properties influence on accuracy and breeder's equation. This extensive list can serve as a check list for the breeders and help to improve prediction by keeping in mind the lessons from the breeder's equation.

Practical steps in linking the original publications to potential breeding process improvement can be detailed as follows:

1. Use of historical data sets to create the training population for GS (publication I, II).
2. Improving selection of traits, which are difficult to measure with low seed availability, such as grain yield (publication I), and quantitatively inherited resistance traits such as FHB resistance in oat (publication II).
3. Shortening of breeding cycle in order to increase genetic gain (publication I).
4. Lowering the costs in phenotyping by predicting missing full sibs in testing environments without considerably affecting prediction accuracy (publication I, III).

5. Testing within year variation of the trial locations in order to quantify the impact of GE interaction in the field testing network (publication III).
6. Showing the impact of GE interaction in the prediction model for overall prediction success (publication III).

5 CONCLUSIONS AND FUTURE PROSPECTS

Crops should feed the ever-growing humankind. The efforts to enhance stable crop production have been substantial, but more advances are needed. As stated in the first chapter, estimated increase in demand of crop production is 50% by the year 2050 (FAO 2017b). Crop production is facing severe challenges such as climate change, limited growing areas, environmentally and public safety driven limitations in the use of fertilizers and pesticides etc. Climate change will increase occurrence of weather extremes, such as droughts, heat-waves and torrential rain (IPCC 2018). Weather extremes will evidently affect crop production. Detailed image was given in Bailes-Serres et al. (2019), where predicted yield losses were described in national level for maize, rice, wheat and soybean (Mills et al. 2018). The detailed stresses included heat, aridity, nutrient stress and flood events. For Nordic region, the primary increasing stress to decrease crop production is heat stress. According to a report from FAO (2017b), 33% of the agriculture land worldwide is degraded moderately or severely. In addition, there are limited options to expand the agriculture land, while changing the land to agriculture would cause heavy environmental, social and economic costs. Therefore, the crop production should be aided by breeding for higher levels of yield per production area. These big challenges can be overcome by appropriate tools, being genomic selection definitively one of these methods to lift the level of yearly yield increase. As we have shown, with genomic selection it is possible to predict demanding traits in the early generations, such as grain yield and FHB resistance traits, improve prediction of the most important trait, grain yield and explore the variability within the testing locations more closely.

Another aiding tool for the future crop improvement is genome editing (Jaganathan et al. 2018). For instance, FHB resistance would present an excellent target for genome editing (II). In wheat, a major QTL, *FHB1* (Li et al. 2019, Su et al. 2019, Rawat et al. 2016), has been found from a Chinese landrace and integrated into modern cultivars. In oat such a major gene has not been found, at least not yet, even when many studies on landraces and exotic sources of resistance have been conducted. Genome editing, although promising, would also be challenging as, a gene should be transferred into a new species, rather than silencing an existing gene.

The next steps for plant researchers and in the future for breeders are to improve the efficiency of photosynthesis, to promote useful microorganism-plant interactions in order to stimulate nutrient intake and to optimize water and nutrient use in the plant (Bailey-Serres et al. 2019). These are by no means easy tasks. Crops do not transform the light energy into carbohydrates as efficient as possible (Zhu et al. 2010), which may be due to that they evolved in water environment where light intensity is less compared to agronomic

conditions (Dann and Leister 2017). Improving photosynthesis is not a new topic, but with modern applications, it might be closer than what we think (Bailey-Serres et al. 2019). Plants by nature engage with microorganism, which improves nutrient intake. By modern agriculture, these benefits have been lost due to use of fertilizers and thus offering a sufficient amount of nutrients for the plants to use. These interactions could be improved especially for phosphate and nitrogen usage. Plants lose water per every carbohydrate that they produce through stomata cells (Bailey-Serres et al. 2019). Responses of stomata cells are not optimal and researchers have been able to improve plant response in water-restricted environments (Hughes et al. 2017, Caine et al. 2019). While these topics are currently under intensive research, the future plant breeders are ready and willing to grab all the available tools to help in producing more food with more environmentally, sociologically and economically stable methods. As have been seen in GS, plant breeders are willing to step into the unknown zone in order to respond to the demands for more food. Plant breeding is not a stable practice but an ever-changing process exploiting multiple methods, which are altered and refreshed by following the latest research.

REFERENCES

- Alqudah, A. M., & Schnurbusch, T. (2017). Heading date is not flowering time in spring barley. *Frontiers in Plant Science*, 8, 896
- Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science*, 322(5903), 881–888
- Anderson, J., Stack, R., Liu, S., Waldron, B., Fjeld, A., Coyne, C., ... Frohberg, R. (2001). DNA markers for Fusarium head blight resistance QTLs in two wheat populations. *Theoretical and Applied Genetics*, 102(8), 1164–1168
- Aranzana, M., Kim, S., Zhao, K., Bakker, E., Horton, M., Jakob, K., ... Nordborg, M. (2005). Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genetics*, 1(5), 0531–0539
- Arbelbide, M., Yu, J., & Bernardo, R. (2006). Power of mixed-model QTL mapping from phenotypic, pedigree and marker data in self-pollinated crops. *Theoretical and Applied Genetics*, 112(5), 876–884
- Arranz-Otaegui, A., Carretero, L., Ramsey, M. N., Fuller, D. Q., & Richter, T. (2018). Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *Proceedings of the National Academy of Sciences*, 115(31), 7925–7930
- Arruda, M. P., Brown, P., Brown-Guedira, G., Krill, A. M., Thurber, C., Merrill, K. R., ... Kolb, F. L. (2016). Genome-wide association mapping of Fusarium head blight resistance in wheat using genotyping-by-sequencing. *The Plant Genome*, 9(1), 1–14
- Asoro, F. G., Newell, M. A., Beavis, W. D., Scott, P. M., Tinker, N. A., & Jannink, J.-L. (2013). Genomic, marker-assisted, and pedigree-BLUP selection methods for β -glucan concentration in elite oat. *Crop Science*, 53, 1894–1906
- Atlin, G. N., Cairns, J. E., & Das, B. (2017). Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Global Food Security*, 12, 31–37
- Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E., & Schroeder, J. I. (2019). Genetic strategies for improving crop yields. *Nature*, 575(7781), 109–118
- Bandeira e Sousa, M., Cuevas, J., de Couto, E., Pérez-Rodríguez, P., Jarquín, D., Fritsche-Neto, R., ... Crossa, J. (2017). Genomic-enabled prediction in maize using kernel models with genotype \times environment interaction. *G3: Genes, Genomes, Genetics*, 7(6), 1995–2014
- Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, 242, 23–36

- Bauer, A. M., Reetz, T. C., & Léon, J. (2006). Estimation of breeding values of inbred lines using best linear unbiased prediction (BLUP) and genetic similarities. *Crop Science*, 46(6), 2685–2691
- Baye, T. M., Abebe, T., & Wilke, R. A. (2011). Genotype-environment interactions and their translational implications. *Personalized Medicine*, 8(1), 59–70
- Bengtsson, T., Consortium, T., Manninen, O., Jahoor, A., & Orabi, J. (2017). Genetic diversity, population structure and linkage disequilibrium in Nordic spring barley (*Hordeum vulgare* L. subsp. *vulgare*). *Genetic Resources and Crop Evolution*, 64(8), 2021–2033
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300
- Bernal-Vasquez, A.-M., Gordillo, A., Schmidt, M., & Piepho, H.-P. (2017). Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. *BMC Genetics*, 18(51)
- Bernardo, Rex. (1994). Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Science*, 34(1), 20
- Bernardo, Rex, & Yu, J. (2007). Prospects for genomewide selection for quantitative traits in maize. *Crop Science*, 47(3), 1082–1090
- Bernardo, R. (2010). Breeding for quantitative traits in plants. 2nd Edition. Stemma Press, Woodbury, Minnesota, US. pp. 4, 73, 177–203, 337
- Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics*, 129(12), 2323–2332
- Beyene, Semagn, Mugo, Tarekegne, Babu, Meisel, ... Crossa, J. (2015). Genetic gains in grain yield through genomic selection in eight bi-parental maize populations under drought stress. *Crop Science*, 55(1), 154–163
- Bowman, J.G.P., Blake, T., LMM, S., Habernicht, D., Daniels, T., & Daniels, J. (1996). Genetic factors controlling digestibility of barley for ruminants. In: *Proceedings, Western Section, American Society of Animal Science* 47.
- Briggs, D. (1978). Barley. Chapman and Hall Ltd., London, UK. pp. 53, 84, 459
- Brim, C. (1966). A modified pedigree method of selection in soybeans. *Crop Science*, 6, 220
- Burgueño, J., Crossa, J., Cotes, J., Vicente, F., & Das, B. (2011). Prediction assessment of linear mixed models for multienvironment trials. *Crop Science*, 51(3), 944–954
- Burgueño, J., de los Campos, G., Weigel, K., & Crossa, J. (2012). Genomic prediction of breeding values when modeling genotype × environment interaction using pedigree and dense molecular markers. *Crop Science*, 52(2), 707–719
- Buerstmayr, H., Ban, T., & Anderson, J. (2009). QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. *Plant Breeding*, 128(1), 1–26

- Büschges, R., Hollricher, K., Panstruga, R., Simons, G., Wolter, M., Frijters, A., ... Schulze-Lefert, P. (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell*, 88(5), 695–705
- Caine, R. S., Yin, X., Sloan, J., Harrison, E. L., Mohammed, U., Fulton, T., ... Gray, J. E. (2019). Rice with reduced stomatal density conserves water and has improved drought tolerance under future climate conditions. *New Phytologist*, 221(1), 371–384
- Cavanagh, C., Morell, M., Mackay, I., & Powell, W. (2008). From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Current Opinion in Plant Biology*, 11(2), 215–221
- Cericola, F., Jahoor, A., Orabi, J., Andersen, J. R., Janss, L. L., & Jensen, J. (2017). Optimizing training population size and genotyping strategy for genomic prediction using association study results and pedigree information. A case of study in advanced wheat breeding lines. *PLOS ONE*, 12, 1–20
- Chaffin, A. S., Huang, Y.-F., Smith, S., Bekele, W. A., Babiker, E., Gnanesh, B. N., ... Tinker, N. A. (2016). A consensus map in cultivated hexaploid oat reveals conserved grass synteny with substantial subgenome rearrangement. *The Plant Genome*, 9(2), 1–21
- Cobb, J. N., Juma, R. U., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., ... Ng, E. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. *Theoretical and Applied Genetics*, 132(3), 627–645
- Coffman, F. (1946). Origin of cultivated oats. *Journal of the American Society of Agronomy*, 38, 983–1002
- Coffman, F. (1977). Oat history, identification and classification. Technical Bulletin No. 1516. Agricultural Research Service United States Department of Agriculture, Washington D.C, US
- Collard, B. C., & Mackill, D. J. (2008). Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1491), 557–572
- Combs, E., & Bernardo, R. (2013). Genomewide selection to introgress semidwarf maize germplasm into U.S. corn belt inbreds. *Crop Science*, 53(4), 1427–1436
- Comstock, R., & Moll, R. (1963). Genotype-environment interactions. National Academy of Sciences, National Research Council, Publ. 982, Washington D.C., 164–196
- Crossa, J., & Cornelius, P. L. (1997). Sites regression and shifted multiplicative model clustering of cultivar trial sites under heterogeneity of error variances. *Crop Science*, 37(2), 406–415
- Crossa, J., Cornelius, P. L., & Yan, W. (2002). Biplots of linear-bilinear models for studying crossover genotype \times environment interaction. *Crop Science*, 42(2), 619–633

- Crossa, J., de los Campos, G., Pérez-Rodríguez, P., Gianola, D., Burgueño, J., Araus, J., ... Braun, H.-J. (2010). Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186(2), 713–724
- Crossa, J., Pérez-Rodríguez, P., de los Campos, G., Mahuku, G., Dreisigacker, S., & Magorokosho, C. (2011). genomic selection and prediction in plant breeding. *Journal of Crop Improvement*, 25(3), 239–261
- Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., ... Varshney, R. K. (2017). Genomic selection in plant breeding: methods, models, and perspectives. *Trends in Plant Science*, 22(11), 961–975
- Cuevas, J., Granato, I., Fritsche-Neto, R., Montesinos-Lopez, O. A., Burgueño, J., Sousa, M., & Crossa, J. (2018). Genomic-enabled prediction kernel models with random intercepts for multi-environment trials. *G3: Genes, Genomes, Genetics*, 8(4), 1347–1362
- Daetwyler, H. D., Villanueva, B., Bijma, P., & Woolliams, J. A. (2007). Inbreeding in genome-wide selection. *Journal of Animal Breeding and Genetics*, 124(6), 369–376
- Dann, M., & Leister, D. (2017). Enhancing (crop) plant photosynthesis by introducing novel genetic diversity. *Philosophical Transactions of the Royal Society B*, 372(1730), 20160380
- da Silva Dias, J.C. (2003). Protocol for broccoli microspore culture. In: Doubled haploid production in crop plants. A manual. Edited by Maluszynski, M., Kasha, K.J., Forster, B.P., Szarejko, I. Kluwer Academic Publishers, Dordrecht, Netherlands. pp 195–204
- Dawson, J. C., Endelman, J. B., Heslot, N., Crossa, J., Poland, J., Dreisigacker, S., ... Jannink, J.-L. (2013). The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Research*, 154, 12–22
- de los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., ... Cotes, J. (2009). Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 182(1), 375–385
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., & Calus, M. P. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics*, 193(2), 327–345
- Devauux, P., & Kasha, K. (2009). Overview of barley doubled haploid production. In: *Advances in haploid production in higher plants*. Edited by Touraev, A., Forster, B.P., Jain, S.M. Springer, Berlin, Germany. pp 47–64
- Doll, H. (1983). Barley seed proteins and possibilities for their improvement. In: *Seed proteins, biochemistry, genetics, nutritive value*. Edited by Gottschalk, W., Müller, H.P. vol 2. Springer, Dordrecht, Netherlands. pp 207–223

- Donnelly, P. (2008). Progress and challenges in genome-wide association studies in humans. *Nature*, 456(7223), 728–731
- Eberhart, S., & Russell, W. (1966). Stability parameters for comparing varieties. *Crop Science*, 6, 36–40
- Eberhart, S. A. (1970). Factors effecting efficiencies of breeding methods. *African soils*, 15, 655–680
- Edwards, S., Buntjer, J. B., Jackson, R., Bentley, A. R., Lage, J., Byrne, E., ... Hickey, J. M. (2019). The effects of training population design on genomic prediction accuracy in wheat. *Theoretical and Applied Genetics*, 132(7), 1943–1952
- Emrich, K., Wilde, F., Miedaner, T., & Piepho, H. (2008). REML approach for adjusting the Fusarium head blight rating to a phenological date in inoculated selection experiments of wheat. *Theoretical and Applied Genetics*, 117(1), 65–73
- Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., ... Goddard, M. E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *Journal of Dairy Science*, 95(7), 4114–4129
- Falconer, D. S., & Mackay, T. F. C. (1996). Introduction to quantitative genetics. 4th edition. Longman Group Ltd., Harlow, UK.
- Food and Agriculture Organization of the United Nations FAO (2017a). FAOSTAT. Accessed 15.12.2019. <http://www.fao.org/faostat/en/#data/QC>
- Food and Agriculture Organization of the United Nations FAO (2017b). The future of food and agriculture – Trends and challenges. Rome.
- Finlay, K., & Wilkinson, G. (1963). The analysis of adaptation in a plant-breeding programme. *Crop and Pasture Science*, 14(6), 742–754
- Flint-Garcia, S. A., Thornsberry, J. M., Edwards, S., & Buckler, E. S. (2003). Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology*, 54(1), 357–374
- Fernandes, S. B., Dias, K. O., Ferreira, D. F., & Brown, P. J. (2018). Efficiency of multi-trait, indirect, and trait-assisted genomic selection for improvement of biomass sorghum. *Theoretical and Applied Genetics*, 131(3), 747–755
- Forster, B. (2014). Accelerated plant breeding. *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, 9(043) 1–16
- Fu, Y.-B. (2018). Oat evolution revealed in the maternal lineages of 25 *Avena* species. *Scientific Reports*, 8(4252), 1–12
- Gauch, H. G. (1988). Model selection and validation for yield trials with interaction. *Biometrics*, 44(3), 705–715

- Gauch, H. G. (2013). A simple protocol for ammi analysis of yield trials. *Crop Science*, 53(5), 1860–1869
- Gaynor, C. R., Gorjanc, G., Bentley, A. R., Ober, E. S., Howell, P., Jackson, R., ... Hickey, J. M. (2017). A two-part strategy for using genomic selection to develop inbred lines. *Crop Science*, 57, 2372–2386
- Gianola, D., Fernando, R. L., & Stella, A. (2006). Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 173(3), 1761–1776
- Gianola, D., de los Campos, G., Hill, W. G., Manfredi, E., & Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics*, 183(1), 347–363
- Gibson, G. (2012). Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13(2), 135–145
- Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298), 369–375
- Gordon, D., & Finch, S. J. (2005). Factors affecting statistical power in the detection of genetic association. *Journal of Clinical Investigation*, 115(6), 1408–1418
- Habier, D., Fernando, R. L., Kizilkaya, K., & Garrick, D. J. (2011). Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics*, 12(186), 1–12
- Hakala, K., Jauhiainen, L., Himanen, S. J., Rötter, R., Salo, T., & Kahiluoto, H. (2012). Sensitivity of barley varieties to weather in Finland. *The Journal of Agricultural Science*, 150(2), 145–160
- Hammer, K. (1984). Das domestikationssyndrom. *Die Kulturpflanze*, 32(1), 11–34
- Harlan, J., & Wet, J. (1971). Toward a rational classification of cultivated plants. *taxon*, 20(4), 509–517
- Hartl, D., & Clark, A. (2007). *Principles of population genetics*. 4th Edition. Sinauer Associates, Inc. Publishers, Sunderland, US. pp 20
- Hughes, J., Hepworth, C., Dutton, C., Dunn, J. A., Hunt, L., Stephens, J., ... Gray, J. E. (2017). Reducing stomatal density in barley improves drought tolerance without impacting on yield. *Plant Physiology*, 174(2), 776–787
- Hästbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A., & Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nature Genetics*, 2(3), 204–211
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., & Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science*, 92(2), 433–443

- Hayes, B., Visscher, P., & Goddard, M. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*, 91(1), 47–60
- Hazel, L. N., Dickerson, G. E., & Freeman, A. E. (1994). the selection index—then, now, and for the future. *Journal of Dairy Science*, 77(10), 3236–3251
- Heffner, E. L., Lorenz, A. J., Jannink, J.-L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: gain per unit time and cost. *Crop Science*, 50
- Henderson, C. (1963). Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding*, NAS-NRC 982, Washington D.C., US, 141–163
- Henderson, C. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 31(2), 423–447
- Heslot, N., Akdemir, D., Sorrells, M. E., & Jannink, J.-L. (2014). Integrating environmental covariates and crop modeling into the genomic selection framework to predict genotype by environment interactions. *Theoretical and Applied Genetics*, 127(2), 463–480
- Heslot, N., Yang, H.-P., Sorrells, M. E., & Jannink, J.-L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Science*, 52(1), 146–160
- Hickey, J., Chiurugwi, T., Mackay, I., Powell, W. & Implementing genomic selection in CGIAR breeding programs workshop participants (2017) Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics*, 49(9), 1297–1302
- Hyten, D. L., Song, Q., Choi, I.-Y., Yoon, M.-S., Specht, J. E., Matukumalli, L. K., ... Cregan, P. B. (2008). High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *Theoretical and Applied Genetics*, 116(7), 945–952
- Intergovernmental Panel on Climate Change IPCC (2018). Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. Edited by Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield. In Press.
- Isidro, J., Jannink, J.-L., Akdemir, D., Poland, J., Heslot, N., & Sorrells, M. E. (2015). Training set optimization under population structure in genomic selection. *Theoretical and Applied Genetics*, 128(1), 145–158
- The International Wheat Genome Sequencing Consortium IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science*, 361(6403)

- Jaganathan, D., Ramasamy, K., Sellamuthu, G., Jayabalan, S. & Venkataraman, G. (2018). CRISPR for crop improvement: an update review. *Frontiers in Plant Science*, 9, 985
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning, with applications in R*. Springer, New York, US. pp 175–194
- Jannink, J.-L., Lorenz, A. J., & Iwata, H. (2010). Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics*, 9(2), 166–177
- Jarquín, D., Crossa, J., Lacaze, X., Cheyron, P., Daucourt, J., Lorgeou, J., ... de los Campos, G. (2014). A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics*, 127(3), 595–607
- Jarquín, D., da Silva, C., Gaynor, C. R., Poland, J., Fritz, A., Howard, R., ... Crossa, J. (2017). Increasing genomic-enabled prediction accuracy by modeling genotype \times environment interactions in Kansas wheat. *The Plant Genome*, 10(2), 1–15
- Jellen, E., Gill, B., & Cox, T. (1994). Genomic in situ hybridization differentiates between A/D- and C-genome chromatin and detects intergenomic translocations in polyploid oat species (genus *Avena*). *Genome*, 37(4), 613–618
- Jellen, E., Rines, H., Fox, S., Davis, D., Phillips, R., & Gill, B. (1997). Characterization of ‘Sun II’ oat monosomics through C-banding and identification of eight new ‘Sun II’ monosomics. *Theoretical and Applied Genetics*, 95(8), 1190–1195
- Jellen, E., & Beard, J. (2000). Geographical distribution of a chromosome 7c and 17 intergenomic translocation in cultivated oat. *Crop Science*, 40, 256–263
- Jensen, J., Mäntysaari, E., Madsen, P., & Thompson, R. (1997). Residual maximum likelihood estimation of (co)variance components in multivariate mixed linear models using average information. *Journal of the Indian Society of Agricultural Statistics*, 49, 215–236
- Jia, Y., & Jannink, J.-L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics*, 192(4), 1513–1522
- Juliana, P., Singh, R. P., Singh, P. K., Crossa, J., Huerta-Espino, J., Lan, C., ... Sorrells, M. E. (2017). Genomic and pedigree-based prediction for leaf, stem, and stripe rust resistance in wheat. *Theoretical and Applied Genetics*, 130(7), 1415–1430
- Jørgensen, I.H. (1992) Discovery, characterization and exploitation of Mlo powdery mildew resistance in barley. *Euphytica*, 63(1-2), 141–152
- Kislev, M. E., Nadel, D., & Carmi, I. (1992). Epipalaeolithic (19,000 BP) cereal and fruit diet at Ohalo II, Sea of Galilee, Israel. *Review of Palaeobotany and Palynology*, 73, 161–166

- Krause, M. R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R. P., ... Mondal, S. (2019). Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat. *G3: Genes, Genomes, Genetics*, 9, 1231–1247
- Kristensen, P. S., Jahoor, A., Andersen, J. R., Cericola, F., Orabi, J., Janss, L. L., & Jensen, J. (2018). Genome-wide association studies and comparison of models and cross-validation strategies for genomic prediction of quality traits in advanced winter wheat breeding lines. *Frontiers in Plant Science*, 9, 69, 1–15
- Kristensen, P., Jahoor, A., Andersen, J., ad Orabi, J., Janss, L., & Jensen, J. (2019). Multi-trait and trait-assisted genomic prediction of winter wheat quality traits using advanced lines from four breeding cycles. *Crop Breeding, Genetics and Genomics*, 1, e1900010
- Lander, E., & Schork, N. (1994). Genetic dissection of complex traits. *Science*, 265(5181), 2037–2048
- Larkin, D.L., Lozada, D.N. & Mason, R.E. (2019). Genomic selection—considerations for successful implementation in wheat breeding programs. *Agronomy*, 9(479), 1–18
- Latta, R. G., Bekele, W. A., Wight, C. P., & Tinker, N. A. (2019). Comparative linkage mapping of diploid, tetraploid, and hexaploid *Avena* species suggests extensive chromosome rearrangement in ancestral diploids. *Scientific Reports*, 9, 12298
- Lee, C., Horsley, R., Manthey, F., & Schwarz, P. (1997). Comparisons of β -glucan content of barley and oat. *Cereal Chemistry*, 74(5), 571–575
- Lenaerts, B., Collard, B., & Demont, M. (2019). Review: Improving global food security through accelerated plant breeding. *Plant Science*, 287, 110207
- Li, C.-D., Rossnagel, B. G., & Scoles, G. J. (2000). Tracing the phylogeny of the hexaploid oat with satellite DNAs. *Crop Science*, 40(6), 1755–1763
- Li, T., Zhang, H., Huang, Y., Su, Z., Deng, Y., Liu, H., ... Zhou, Y. (2019). Effects of the *Fhb1* gene on Fusarium head blight resistance and agronomic traits of winter wheat. *The Crop Journal* 7(6), 799–808
- Lin, C., Binns, M., & Lefkovitch, L. (1986). Stability analysis: where do we stand? *Crop Science*, 26(5), 894–900
- Long, A., & Langley, C. (1999). The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. *Genome Research*, 9(8), 720–731
- Longin, F. C., Mi, X., & Würschum, T. (2015). Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theoretical and Applied Genetics*, 128(7), 1297–1306
- Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., ... Jannink, J.-L. (2011). Chapter 2: Genomic selection in plant breeding: knowledge and prospects. *Advances in Agronomy*, 110, 77–123

- Lorenz, A. J., & Smith, K. P. (2015). Adding genetically distant individuals to training populations reduces genomic prediction accuracy in barley. *Crop Science*, 55, 2657–2667
- Lu, Q., Lillemo, M., Skinnens, H., He, X., Shi, J., Ji, F., ... Bjørnstad, Å. (2013). Anther extrusion and plant height are associated with Type I resistance to *Fusarium* head blight in bread wheat line 'Shanghai-3/Catbird.' *Theoretical and Applied Genetics*, 126(2), 317–334
- Lush, J. (1943). *Animal Breeding Plans*. 2nd edition. The Iowa State College Press
- Ly, Hamblin, Rabbi, Melaku, Bakare, Gauch, G., ... Jannink. (2013). Relatedness and genotype × environment interaction affect prediction accuracies in genomic selection: a study in cassava. *Crop Science*, 53(4), 1312–1325
- Lynch, M., & Walsh, B. (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc. Massachusetts, US pp. 379
- Maltecca, C., Parker, K., & Cassady, J. (2012). Application of multiple shrinkage methods to genomic predictions. *Journal of Animal Science*, 1777–1787
- Malysheva-Otto, L. V., Ganai, M. W., & Röder, M. S. (2006). Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genetics*, 7(6), 1–14
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747–753
- Marshall, A., Cowan, S., Edwards, S., Griffiths, I., Howarth, C., Langdon, T., & White, E. (2013). Crops that feed the world 9. Oats- a cereal crop for human and livestock feed with industrial applications. *Food Security*, 5(1), 13–33
- Marttinen, P., & Corander, J. (2010). Efficient Bayesian approach for multilocus association mapping including gene-gene interactions. *BMC Bioinformatics*, 11(443), 1–20
- Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S. O., Wicker, T., ... Stein, N. (2017). A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 544(7651), 427–433
- Massman, J. M., Jung, H.-J. G., & Bernardo, R. (2013). Genomewide selection versus marker-assisted recurrent selection to improve grain yield and stover-quality traits for cellulosic ethanol in maize. *Crop Science*, 53, 1–9
- Mayer, K. F., Waugh, R., Langridge, P., Close, T. J., Wise, R. P., Graner, A., ... Stein, N. (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491(7426), 711–716
- Meuwissen, T., Hayes, B., & Goddard, M. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4), 1819–1829

- Meuwissen, T. H. (2009). Accuracy of breeding values of “unrelated” individuals predicted by dense SNP genotyping. *Genetics Selection Evolution*, 41(35), 1–9
- Michel, S., Ametz, C., Gungor, H., Epure, D., Grausgruber, H., Löschenberger, F., & Buerstmayr, H. (2016). Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theoretical and Applied Genetics*, 129(6), 1179–1189
- Michel, S., Kummer, C., Gallee, M., Hellinger, J., Ametz, C., Akgöl, B., ... Buerstmayr, H. (2018). Improving the baking quality of bread wheat by genomic selection in early generations. *Theoretical and Applied Genetics*, 131(2), 477–493
- Mickelbart, M. V., Hasegawa, P. M., & Bailey-Serres, J. (2015). Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nature Reviews Genetics*, 16(4), 237–251
- Mitchell Fetch, J.W., Tekauz, A., Brown, P.D., Ames, N., Chong, J., Fetch, T.G., ... Green, D.A. (2013) Stride oat. *Canadian journal of Plant Science*, 93(4), 749–753
- Mills, G. Sharps, K., Simpson, D., Pleijel, H., Frei, M., Burkey, K. ... Agrawal, M. (2018). Closing the global ozone yield gap: quantification and cobenefits for multistress tolerance. *Global Change Biology*, (24)4, 869–925
- Muir, W. (2007). Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. *Journal of Animal Breeding and Genetics*, 124(6), 342–355
- Munkvold, J. D., Tanaka, J., Benscher, D., & Sorrells, M. E. (2009). Mapping quantitative trait loci for preharvest sprouting resistance in white wheat. *Theoretical and Applied Genetics*, 119(7), 1223–1235
- Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. S. (2009). Association mapping: critical considerations shift from genotyping to experimental design. *The Plant Cell*, 21(8), 2194–2202
- Naidoo, N., Pawitan, Y., Soong, R., Cooper, D. N., & Ku, C.-S. (2011). Human genetics and genomics a decade after the release of the draft sequence of the human genome. *Human Genomics*, 5(6), 577–622
- Newman, R., Newman, C., & Graham, H. (1989). The hypocholesterolemic function of barley B-glucan. *Cereal Foods World*, 34, 883–886
- Newton, A., Flavell, A. J., George, T. S., Leat, P., Mullholland, B., Ramsay, L., ... Bingham, I. J. (2011). Crops that feed the world 4. Barley: a resilient crop? Strengths and weaknesses in the context of food security. *Food Security*, 3, 141–178
- Nielsen, N., Jahoor, A., Jensen, J., Orabi, J., Cericola, F., Edriss, V., & Jensen, J. (2016). Genomic prediction of seed quality traits using advanced barley breeding lines. *PLOS ONE*, 11(10), e0164494
- Nordborg, M., & Weigel, D. (2008). Next-generation genetics in plants. *Nature*, 456(7223), 720–723

- Norman, A., Taylor, J., Edwards, J., & Kuchel, H. (2018). Optimising genomic selection in wheat: effect of marker density, population size and population structure on prediction accuracy. *G3: Genes, Genomes, Genetics*, 8, 2889–2899
- Oakey, H., Verbyla, A., Pitchford, W., Cullis, B., & Kuchel, H. (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics*, 113(5), 809–819
- Öfversten, J., Jauhiainen, L., & Kangas, A. (2004). Contribution of new varieties to cereal yields in Finland between 1973 and 2003. *The Journal of Agricultural Science*, 142(3), 281–287
- Oliver, R. E., Lazo, G. R., Lutz, J. D., Rubenfield, M. J., Tinker, N. A., Anderson, J. M., ... Jackson, E. W. (2011). Model SNP development for complex genomes based on hexaploid oat using high-throughput 454 sequencing technology. *BMC Genomics*, 12(77), 1–15
- Oliver, R. E., Tinker, N. A., Lazo, G. R., Chao, S., Jellen, E. N., Carson, M. L., ... Jackson, E. W. (2013). SNP discovery and chromosome anchoring provide the first physically-anchored hexaploid oat map and reveal synteny with model species. *PLOS ONE*, 8(3), e58068
- Ouyang, Z., Mowers, R., Jensen, A., Wang, S., & Zheng, S. (1995). Cluster analysis for genotype \times environment interaction with unbalanced data. *Crop Science*, 35(5), 1300–1305
- Panter, D., & Allen, F. (1995a). Using best linear unbiased predictions to enhance breeding for yield in soybean: i. choosing parents. *Crop Science*, 35(2), 397–405
- Panter, D., & Allen, F. (1995b). Using best linear unbiased predictions to enhance breeding for yield in soybean: ii. selection of superior crosses from a limited number of yield trials. *Crop Science*, 35(2), 405–410
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), 2074–2093
- Paulitz, T. C., & Steffenson, B. J. (2011). Biotic stress in barley: disease problems and solutions. In: *Barley*. Edited by Ullrich, S.E. Blackwell Publishing Ltd. New Jersey, US. pp 307–354
- Pauw, D. R., Faris, D., & Willia, C. (1981). Genotype-environment interaction of yield in cereal crops in Northwestern Canada. *Canadian Journal of Plant Science*, 61(2), 255–263
- Peng, Y.-Y., Wei, Y.-M., Baum, B. R., Jiang, Q.-T., Lan, X.-J., Dai, S.-F., & Zheng, Y.-L. (2010). Phylogenetic investigation of *Avena* diploid species and the maternal genome donor of *Avena* polyploids. *Taxon*, 59(5), 1472–1482
- Pérez-Rodríguez, P., J. Crossa, K. Bondalapati, G. Meyer, F. Pita, et al. (2015). A pedigree-based reaction norm model for prediction of cotton yield in multienvironment trials. *Crop Science* 55(3): 1143–1151
- Peterson, D. M. (2001). Oat antioxidants. *Journal of Cereal Science*, 33(2), 115–129

- Piepho, H., Möhring, J., Melchinger, A., & Büchse, A. (2008). BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica*, 161(1–2), 209–228
- Pikkuhookana, P., & Sillanpää, M. (2014). Combined linkage disequilibrium and linkage mapping: Bayesian multilocus approach. *Heredity*, 112(3), 351–360
- Poland, J. A., & Rife, T. W. (2012). Genotyping-by-sequencing for plant breeding and genetics. *The Plant Genome Journal*, 5(3), 92–102
- Porter, J., Xie, L., Challinor, A., Cochrane, K., Howden, Iqbal, M., ... Travasso, M. (2014). Climate change 2014 impacts, adaptation, and vulnerability, Part A: global and sectoral aspects. Cambridge University Press, New York, US, pp 488–504
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909
- Price, A. L., Zaitlen, N. A., Reich, D., & Patterson, N. (2010). New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 11(7), 459–463
- Pritchard, J., ephens, & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959
- Purugganan, M. D., & Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature*, 457(7231), 843–848
- Rafalski, A. (2002). Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, 5(2), 94–100
- Rajhathy, T. & Thomas, H. (1974). Cytogenetics of oats (*Avena* L.). Edited by Wilkes, A., Miscellaneous publications of the genetics society of Canada, 2, pp 67. The genetics society of Canada, Ottawa, Canada.
- Rawat, N., Pumphrey, M.O., Liu, S., Zhang, X., Tiwari, V.K., Ando, K., ... Gill, B.S. (2016). Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore802 forming toxin-like domain conferring resistance to Fusarium head blight. *Nature Genetics* 48, 1576–1580
- Remington, D., Thornsberry, J., Matsuoka, Y., Wilson, L., Whitt, S., Doebley, J., ... Buckler, E. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences*, 98(20), 11479–11484
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281), 1516–1517
- Robertsen, C. D., Hjortshøj, R. L., & Janss, L. L. (2019). Genomic selection in cereal breeding. *Agronomy*, 9(95), 1–16
- Robertson, A. (1959). The sampling variance of the genetic correlation coefficient. *Biometrics*, 15(3), 469–485
- Rutkoski, J, Singh, R., Huerta-Espino, J., Bhavani, S., Poland, J., Jannink, J., & Sorrells, M. (2015). Genetic gain from phenotypic and genomic selection

- for quantitative resistance to stem rust of wheat. *The Plant Genome*, 8(2), 1–10
- Rutkoski, Jessica, Poland, J., Mondal, S., Autrique, E., Pérez, L., Crossa, J., ... Singh, R. (2016). Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. *G3: Genes|Genomes|Genetics*, 6(9), 2799–2808
- Sannemann, W., Huang, B., Mathew, B., & Léon, J. (2015). Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Molecular Breeding*, 35(86), 1–16
- Schaeffer, L. R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123(4), 218–223
- Schmidt, M., Kollers, S., Maasberg-Prelle, A., Großer, J., Schinkel, B., Tomerius, A., ... Korzun, V. (2016). Prediction of malting quality traits in barley based on genome-wide marker data to assess the potential of genomic selection. *Theoretical and Applied Genetics*, 129(2), 203–213
- Schulthess, A., Wang, Y., Miedaner, T., Wilde, P., Reif, J. C., & Zhao, Y. (2016). Multiple-trait- and selection indices-genomic predictions for grain yield and protein content in rye for feeding purposes. *Theoretical and Applied Genetics*, 129(2), 273–287
- Schulthess, A. W., Zhao, Y., Longin, F. C., & Reif, J. C. (2018). Advantages and limitations of multiple-trait genomic prediction for Fusarium head blight severity in hybrid wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics*, 131(3), 685–701
- Simons, M. D. (1985). Crown rust. In: *The Cereal Rusts, Diseases, Distribution, Epidemiology, and Control*. Vol. 2. Edited by Roelfs, A.P. & Bushnell, W.R. Academic Press Inc. Ltd. London, UK. pp 131–172
- Su, Z., Bernardo, A., Tian, B., Chen, H., Wang, S., Ma, H., ... Bai, G. (2019). A deletion mutation in *TaHRC* confers *Fhb1* resistance to Fusarium head blight in wheat. *Nature Genetics* 51, 1099–1105
- Thomas, H. (1995). Oats. In: *Evolution of crop plants*. 2nd edition. Edited by Smartt, J., Simmonds, N.W. Longman Scientific & Technical, Singapore. pp 132–137
- Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., & Edward, B. S. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. *Nature Genetics*, 28(3), 286–289
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288
- Valentine, J., Cowan, A., & Marshall, A. (2011). Oat breeding. In: *Oats chemistry and technology*. 2nd edition. Edited by Webster, F.H. & Wood, P.J. AACC International, Inc. St. Paul, Minnesota, US. pp 11–30

- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423
- VanRaden, P. M., Tassell, C. P., Wiggans, G. R., Sonstegard, T. S., Schnabel, R. D., Taylor, J. F., & Schenkel, F. S. (2009). Invited Review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92(1), 16–24
- Verbyla, K. L., Hayes, B. J., Bowman, I. J., & Goddard, M. E. (2009). Accuracy of genomic selection using stochastic search variable selection in Australian Holstein Friesian dairy cattle. *Genetics Research*, 91(5), 307–311
- Voss-Fels, K. P., Stahl, A., & Hickey, L. T. (2019). Q&A: modern crop breeding for future food security. *BMC Biology*, 17(18), 1–7
- Welch, R. (2011). Nutrient composition and nutritional quality of oats and comparisons with other cereals. In: *Oats chemistry and technology*. 2nd edition. Edited by Webster, F.H. & Wood, P.J. AACC International, Inc. St. Paul, Minnesota, US. pp 95–107
- Wendorf, F., Schild, R., Hadidi, N., Close, A. E., Kobusiewicz, M., Wieckowska, H., ... Haas, H. (1979). Use of barley in the Egyptian late paleolithic. *Science*, 205(4413), 1341–1347
- Whittaker, J. C., Thon, R., & Denham, M. C. (2000). Marker-assisted selection using ridge regression. *Genetics Research*, 75(2), 249–252
- Williams, E. (1952). The interpretation of interactions in factorial experiments. *Biometrika*, 39(1/2), 65–81
- Wold, H. (1985). Factors influencing the outcome of economic sanctions. *Trabajos de Estadística Y de Investigación Operativa*, 36(3), 325–338
- Thomas, W.T.B., Forster, B., & Gertsson, B. (2003). Doubled haploids in breeding. In: *Doubled haploid production in crop plants. A manual*. Edited by Maluszynski, M., Kasha, K.J., Forster, B.P., Szarejko, I. Kluwer Academic Publishers, Dordrecht, Netherlands. pp 333–349
- Yan, H., Martin, S. L., Bekele, W. A., Latta, R. G., Diederichsen, A., Peng, Y., & Tinker, N. A. (2016). Genome size variation in the genus *Avena*. *Genome*, 59(3), 209–220
- Yan, W., Frégeau-Reid, J., Pageau, D., & Martin, R. (2016). Genotype-by-environment interaction and trait associations in two genetic populations of oat. *Crop Science*, 56, 1136–1145
- Yates, F., & Cochran, W. (1938). The analysis of groups of experiments. *The Journal of Agricultural Science*, 28(4), 556–580
- Yu, J., Arbelhide, M., & Bernardo, R. (2005). Power of in silico QTL mapping from phenotypic, pedigree, and marker data in a hybrid breeding program. *Theoretical and Applied Genetics*, 110(6), 1061–1067
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I., Yamasaki, M., Doebley, J. F., ... Buckler, E. S. (2006). A unified mixed-model method for association

- mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2), 203–208
- Yu, J. Holland, J. B., McMullen, M. D., & Buckler, E. S. (2007). Genetic design and statistical power of nested association mapping in maize. *Genetics*, 178(1), 539–551
- Zhao, K., Aranzana, M., Kim, S., Lister, C., Shindo, C., Tang, C., ... Nordborg, M. (2007). An *Arabidopsis* example of association mapping in structured samples. *PLoS Genetics*, 3(1), 0071–0082
- Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and prospects of association mapping in plants. *The Plant Genome*, 1(1), 5–20
- Zhu, X.-G., Long, S. P., & Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Plant Biology*, 61(1), 235–261
- Zohary, D., Hopf, M., & Weiss, E. (2012). *Domestication of Plants in the Old World*. 4th edition. pp. 22, 52, 58